

Rowan University

Rowan Digital Works

Theses and Dissertations

9-11-2019

Examining teacher perceptions of the Stronge Teacher Effectiveness Performance Evaluation System

Tiffany A. Lynch
Rowan University

Follow this and additional works at: <https://rdw.rowan.edu/etd>



Part of the [Educational Administration and Supervision Commons](#)

Let us know how access to this document benefits you -
share your thoughts on our feedback form.

Recommended Citation

Lynch, Tiffany A., "Examining teacher perceptions of the Stronge Teacher Effectiveness Performance Evaluation System" (2019). *Theses and Dissertations*. 2735.
<https://rdw.rowan.edu/etd/2735>

This Dissertation is brought to you for free and open access by Rowan Digital Works. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of Rowan Digital Works. For more information, please contact LibraryTheses@rowan.edu.

**EXAMINING TEACHER PERCEPTIONS OF THE
STRONGE TEACHER EFFECTIVENESS
PERFORMANCE EVALUATION SYSTEM**

by

Tiffany A. Lynch

A Dissertation

Submitted to the
Department of Educational Services and Leadership
College of Education
In partial fulfillment of the requirement
For the degree of
Doctor of Education
at
Rowan University
August 26, 2019

Dissertation Chair: James Coaxum III, Ph.D.

Dedications

This dissertation is dedicated to my parents for everything they did to make this accomplishment possible. To my mom and dad, thank you for expecting more of me even when I wasn't sure it was possible, and thank you for all of the guidance you have given me. This would never have been possible without your support and encouragement during my life. For as long as I can remember, you instilled in me the importance of an education and for that I will always be grateful.

To my husband, Ben, thank you for being my sounding board and my punching bag all at the same time. I would never have accomplished this without you. You made sure that I reached my goal, even when I didn't believe it was possible.

To Charlotte, thank you for being my inspiration and motivation. I want you to know that you can accomplish anything you set your mind to. Someday you will reach for the stars and touch them! I will make sure of it!

Acknowledgements

I would like to both acknowledge and thank my dissertation chair, Dr. James Coaxum. You stepped in to help me when you didn't have to and for that I cannot thank you enough. For your guidance and support during this process, I am forever grateful. I would also like to thank the other members of my committee, Dr. JoAnn Manning and Dr. Cary Fields, for their expertise and knowledge during this process. Your assistance was invaluable.

To Leslie, you were a constant source of friendship, support, and help during this process. You listened to me every time I needed someone and stopped me from quitting when that was my answer. Your wisdom and counsel during this were priceless. Thank you!

Abstract

Tiffany A. Lynch
EXAMINING TEACHER PERCEPTIONS OF THE
STRONGE TEACHER EFFECTIVENESS
PERFORMANCE EVALUATION SYSTEM
2018 - 2019
James Coaxum III, Ph.D.
Doctor of Education

The purpose of this convergent, parallel mixed methods research study was to gather and analyze data from teachers who have been observed and evaluated using the Stronge Teacher Effectiveness Performance Evaluation System, with regards to it being able to create positive change in both instruction and student achievement, as well as the quality of feedback received on observations and summative evaluations. Data were collected using three instruments: a survey, interviews, and an analysis of summative evaluation scores, and the data led to six emerging themes. The participants did not perceive that the system had the potential to bring about change, in either instruction or student achievement. Participants perceived that administrator use of the system was for compliance purposes rather than a true desire to improve education in the district and that the supervision of teachers needed to shift from completely evaluative to a focus on instructional supervision and professional development. The study found that there was a need to create a connection between observations, evaluations, and professional growth opportunities.

Table of Contents

Abstract	v
List of Tables	x
Chapter 1: Introduction to the Study.....	1
Recent Teacher Evaluation Reform Efforts in New Jersey	5
TeachNJ	6
AchieveNJ.....	7
The Stronge Teacher Effectiveness Performance Evaluation System	8
The Danielson Framework.....	8
The Marzano Teacher Evaluation Model	9
City Public Schools and the Stronge TEPES	10
Statement of the Problem.....	12
Purpose of the Study	13
Significance of the Study	14
Practice.....	15
Policy	16
Theoretical Framework.....	16
Definition of Terms.....	19
Conclusion	22
Chapter 2: Literature Review	23
History of Teacher Evaluations in the United States	25

Table of Contents (Continued)

History of Teacher Evaluations in New Jersey	27
Reform Efforts	29
No Child Left Behind	29
Race to the Top	30
Common Core State Standards	30
PARCC	31
Teacher Evaluation Reform	32
Teacher Point of View on Evaluation.....	32
Administrator Point of View on Evaluation	37
Supervision of Teachers.....	40
Conclusion	43
Chapter 3: Methodology	44
Research Design: Mixed Methods	48
Research Design: Approach	49
Sampling Strategy	50
Participants.....	51
Setting	52
Role of Researcher: Worldview	53
Instrumentation and Data Collection	53
Survey	54
Interview Questions	55
Summative Evaluation Scores	56
Analysis of Data.....	57
Triangulation of Data.....	58
Conclusion	58
Chapter 4: Findings	60
Description of the Design	63

Table of Contents (Continued)

Description of Participants.....	64
Quantitative Findings	67
Overall Quality and Impact	68
Perception of Implementation	70
Quality of Feedback	73
The Change in Summative Evaluation Scores	83
Qualitative Findings	88
Conflicting Perceptions of the Implementation of the Stronge TEPES.....	91
Successful Implementation.....	91
A Challenging Implementation	94
The Potential for Change	96
Increase in Paperwork with a Decrease in Quality Instruction	97
Increase in Accountability for Instruction	99
Administrator Use of Cookie Cutter Feedback.....	102
Instruction Unaffected by the Stronge TEPES	106
Lack of Influence on Professional Development	109
Synthesis of Findings.....	111
Conclusion	113
Chapter 5: Discussion and Implications	115
Discussion of Findings.....	117
Perceptions of the Implementation of Stronge TEPES	118
The Potential to Bring About Change	120
Instruction Unaffected by the Stronge TEPES	121
Administrator Use of Cookie Cutter Feedback	123
The Change in Summative Evaluation Scores	125
Lack of Influence on Professional Development	126
Research Questions	127

Table of Contents (Continued)

Research Question One	127
Research Question Two	128
Research Question Three	128
Research Question Four	128
Reflection on Theoretical Framework	129
Leadership	132
Recommendations	133
Implications	134
Practice	134
Policy	135
Future Research	137
Limitations	138
Trustworthiness	139
Confirmability	139
Credibility	140
Dependability	141
Transferability	141
Conclusion	141
References	144
Appendix A: Survey Questions	150
Appendix B: Interview Protocol	152

List of Tables

Table	Page
Table 1. Survey Participants’ Total Years Teaching in City Public Schools	66
Table 2. Survey Participants’ Grade Level Teaching Assignment	67
Table 3. Overall Quality of the Evaluation Process, including Observations	68
Table 4. Overall Impact of the Stronge TEPES on Professional Practices	69
Table 5. Overall Quality of the Implementation of the Stronge TEPES	71
Table 6. The Ability to Create Positive Change in the District	72
Table 7. Teacher Evaluation Improves the Learning of my Students	73
Table 8. Written Feedback Helped Identify Strengths in Teaching	74
Table 9. Written Feedback Helped Identify Areas Needing Improvement in Teaching ..	75
Table 10. Written Feedback after Formal Observations was Valuable	76
Table 11. Written Feedback after Formal Observations was Meaningful, Identifying Teaching Practices	77
Table 12. Written Feedback after Formal Observations led to Learning about Teaching	79
Table 13. Written Feedback after Formal Observations had Positive Influence on Teaching Practices	80
Table 14. Written Feedback after Formal Observations caused Reflection on Teaching	81
Table 15. Written Feedback after Formal Observations has led to Positive Changes in Teaching Practices	82
Table 16. Annual Evaluation leads to Professional Development and Support for Improvement	83
Table 17. Summative Evaluation Scores	85
Table 18. Interview Participants	90

Chapter 1

Introduction to the Study

On February 17, 2009, the American Recovery and Reinvestment Act (ARRA) was signed into law by President Barack Obama (Race to the Top, 2009). The ARRA created a competitive grant known as Race to the Top (RTT), which was available to individual states and totaled 4.35 billion dollars. The funding was meant to create innovation and reform in local K-12 districts in each individual, qualifying state. Each state was able to submit an application and was awarded points based on their inclusion of four criteria: the adoption of standards and assessments to prepare students to be college and career ready, the creation of a data system that measured student growth and success, the implementation of a system to recruit, develop, reward, and/or retain effective teachers and principals, and the execution of a plan to turn around the lowest performing schools (Race to the Top, 2009).

The RTT grant (2009) sparked ongoing conversation in education about teacher evaluation and the supervision of teachers because there are differing opinions about both. Recently educators across the state of New Jersey felt the impact of changes to teacher evaluation through the mandates of AchieveNJ (2013), which are a result of the federal Race to the Top Grant (RTT) and the TeachNJ law (TeachNJ, 2012). Under the requirements of RTT (2009), an improved teacher evaluation system was included as one of the ways for a state to obtain federal funding for implementation and adoption purposes (Race to the Top, 2009). Under Governor Chris Christie's administration, the state of New Jersey agreed to this specific requirement. RTT (2009) sparked controversy throughout both the country and the state of New Jersey because it shifted the ways in

which both educational practices and protocols have typically transpired, including creating a direct correlation between student achievement and teacher performance.

Additionally, Race to the Top (2009) allowed districts to receive federal funding by including a performance-based pay system, which is in direct contrast to the views of both national and state education unions. Ramirez (2010) reminds us of the dangers that can exist with a performance-based pay system. In a classroom, there are factors such as a teacher's ability to choose his/her students that are beyond a teacher's control. Without having any control over these factors, teachers are affected by students' attendance, students' nutrition, students' early childhood educational experiences, and family resources, all of which are beyond a teacher's control (Ramirez, 2010). Each class and every student presents its own set of problems. This type of policy can also have a negative impact on the morale of teachers because of the inherent competition it creates. Finally, it can lead to many administrative issues such as an increase in both paperwork and grievances that may occur (Ramirez, 2010).

Ramirez's (2010) analysis demonstrates the need for teachers to be evaluated on several factors rather than just student performance on a standardized test. Teacher evaluation should include a preponderance of evidence from a variety of sources including classroom instruction, self-evaluation completed by a teacher, and feedback from all of the stakeholders in the school community: students, parents, colleagues, and administrators. Without other factors calculated in, there might be a sharp decline in teachers willing to work with students who are most at-risk both behaviorally and academically.

In the state of New Jersey, districts were required to make very specific changes to previous teacher evaluation systems because of the AchieveNJ (2013) mandates. One of these changes was that each district in the state had to choose an evaluation model which included a four-point rating scale for a teacher's summative evaluation. Prior to this, TeachNJ (2012), the tenure law in the state, regulated teacher observation and evaluation. Observations consisted mostly of a narrative description of a lesson accompanied by qualitative comments from a single observer, who may have been a supervisor, assistant principal, or principal. The same was true of the annual summative evaluations that teachers received. Prior to the 2012 amendment of Title 18A, Education, Section 18A:28-5 (TeachNJ, 2012), the only requirements for observation of teachers related to the number of times a teacher was observed. Non-tenured teachers were observed three times a year, and tenured teachers were observed once a year. The amendment to TeachNJ (2012) included specific policy changes, including the addition of quantitative measures to something that was previously completely qualitative allowing administrators, members of Boards of Education, and perhaps even politicians to measure the worth of a teacher by an evaluation number.

Former Governor Christie, in compliance with Race to the Top (2009), also mandated the inclusion of student achievement in each teacher's summative evaluation. This aligned to the Race to the Top (2009) criteria of the creation of a data system that measured student growth and success. A teacher's summative evaluation is based on a score, between zero and four, and combines teacher practice, student growth objectives, (SGOs), and student growth percentiles (SGPs) (AchieveNJ, 2013).

Teacher practice, according to AchieveNJ (2013), is the evidence collected by an observer during a minimum number of classroom observations: three for non-tenured teachers and two for tenured teachers. The specific number of observations required for each teacher is regulated by TeachNJ (2012). Non-tenured teachers have to be observed a minimum of one time during the first semester and one time during the second semester (TeachNJ, 2012). During observations, evaluators collect evidence that is subsequently used as a measure of teacher effectiveness (AchieveNJ, 2013).

Student achievement was included through two measures. The first measure, SGPs, are formulated by the New Jersey Department of Education (NJ DOE) using student growth on a standardized test such as New Jersey Assessment of Skills and Knowledge (NJASK), the Partnership for Assessment of Readiness for College and Careers (PARCC) (PARCC, 2016), or the New Jersey Student Learning Assessment (NJSLA) (Spring Assessment, 2018). New Jersey measures growth for each individual student by comparing the change in his or her achievement on the state standardized assessment from one year to the student's "academic peers" (all other students in the state who had similar historical test results). This comparative change in achievement is reported on a 1 to 99 scale. A teacher receives an SGP from the median score of all of his/her students (AchieveNJ, 2013). The second measure, SGOs, are objectives or goals meant to show a teacher's ability to guide students' growth through classroom instruction, performance on benchmark assessments, and/or pre/post assessments (AchieveNJ, 2013).

AchieveNJ (2013) targets teachers of grades four through eight who teach either English Language Arts (ELA) and/or Mathematics as tested teachers. These teachers

each receive an SGP. All other teachers are referred to as non-tested teachers through this mandate (AchieveNJ, 2013). Non-tested teachers are required to create two SGOs using pre-assessment data, and tested teachers are required to create one SGO using pre-assessment data, as well as receiving one SGP through state assessment data. For non-tested teachers, a summative evaluation is weighted 85% of teacher practice and 15% of SGO results. For tested teachers, a summative evaluation is weighted 70% of teacher practice, 5% for SGPs, and 25% for SGOs (AchieveNJ, 2013). Because of these initiatives, it is essential to determine their impact, as well as the perception of the teachers who are being evaluated using these systems in order to inform future practice for both administrators completing observations and evaluations as well as teachers in the areas of professional growth and learning.

Recent Teacher Evaluation Reform Efforts in New Jersey

The education system in the state of New Jersey saw drastic reform efforts under the Governor Christie administration. New standards were introduced and adopted by districts; this was a direct result of the state of New Jersey receiving funds through the Race to the Top Grant (2009) and implementing the mandated changes dictated by the grant. New standardized tests replaced the basic skills assessments that previously existed, and teacher evaluation saw a complete overhaul under both TeachNJ (2012) and AchieveNJ (2013). Each of these reform efforts and changes correlated to one of the criteria of Race to the Top (2009): the adoption of standards and assessments to prepare students to be college and career ready, the creation of a data system that measured student growth and success, and the implementation of a system to recruit, develop, reward, and/or retain effective teachers and principals.

TeachNJ. In August 2012, New Jersey amended the TeachNJ Act, a law regarding tenure for teachers, vice/assistant principals, and principals in the state. TeachNJ (2012) requires an educator to serve four years and one day of employment before tenure is awarded. Additionally, to obtain tenure, every new teacher in the state must complete one year of mentoring during their first year of employment, as well as receive a rating of highly effective or effective in two of the three following years. Principals, vice principals, and assistant principals must receive a summative rating of effective or highly effective at the completion of school years during the first three years of employment (TeachNJ, 2012).

TeachNJ (2012) also speaks directly to the ways in which teachers, principals, vice principals, and assistant principals can lose tenure. If a summative rating of ineffective is followed by a summative rating of ineffective the following year, a superintendent must file tenure charges. Similarly, if a summative rating of partially effective is followed by a summative rating of ineffective the following year, a superintendent must file tenure charges. In a case of a summative rating of ineffective following by a summative rating of partially effective the following year or a summative rating of partially effective two years in a row, a superintendent has the option of filing tenure charges or deferring to a third year. If in the third year, a summative rating of partially effective or ineffective occurs, then tenure charges must be filed (AchieveNJ, 2013). Therefore, either two or three consecutive years of either a summative evaluation score of ineffective or partially ineffective ratings become cause for termination, even of a tenured teacher, principal, or vice principal.

AchieveNJ. Under Governor Christie’s administration, public school districts in the state of New Jersey were required to adhere to the mandate of AchieveNJ (2013), the details and support structures necessary to carry out the TeachNJ (2012) law. TeachNJ (2012) can be described as the law pertaining to evaluation and supervision, while AchieveNJ (2013) encompasses all of the requirements outlined by the NJ DOE in order for districts to carry out the law.

AchieveNJ (2013) required each district to choose a teacher evaluation model that included both teacher practice and student achievement. Initially, the NJ DOE presented several models for districts to choose from, including The Danielson Framework (2007), The Stronge Teacher Effectiveness Performance Evaluation System (TEPES) (2006), and The Marzano Teacher Evaluation Model (2013), the three most commonly selected during the implementation process. Each of these three evaluation systems included a four-point rating scale, which was one of the primary requirements of AchieveNJ (2013). At that time, it was also governed that each teacher be observed a minimum of three times a year, with the length dependent on their tenure status. Short observations were a minimum of 20 minutes, and long observations were a minimum of 40 minutes or one class period. It was required to complete two long observations for a non-tenured teacher and one long observation for a tenured teacher. In a recent change, tenured teachers need to only be observed two times a year, while non-tenured teachers maintain the requirement of three observations each year due to a deluge of responsibilities and impossibility of completing the vast majority of observations. Each teacher is also required to create SGOs, one for tested teachers and two for non-tested teachers. Finally,

the teachers who taught tested subjects, ELA or Mathematics in grades 4-8, would also receive a SGP.

The Stronge Teacher Effectiveness Performance Evaluation System. Dr. James Stronge, the Heritage Professor in the Educational Policy, Planning, and Leadership Department at the College of William and Mary, developed a system of teacher evaluation to collect and present data to document teacher performance based on well-defined job expectations (Stronge, 2015). Having received his doctorate in the area of educational administration and planning from the University of Alabama, Stronge has worked as a teacher, counselor, and district level administrator. With research interests in policy and practice related to teacher quality, Stronge's consulting work focuses on how to identify effective teachers and how to increase teacher effectiveness (Stronge, 2006).

The Stronge Teacher Effectiveness Performance Evaluation System (TEPES) (2006), one of the NJ DOE approved systems, includes seven standards: professional knowledge, instructional planning, instructional delivery, assessment of/for learning, learning environment, professionalism, and student achievement. This differs from the other evaluation tools as the Stronge TEPES (2006) does not include subcategories (elements or components) for each of the performance standards. Instead, there is a list of performance indicators that evaluators can use as a guide for an overall summative rating in each standard. These performance indicators are not used as a checklist. The four-point rating scale for this evaluation tool includes: ineffective, partially effective, effective, and highly effective.

The Danielson Framework. Charlotte Danielson (Danielson, 2007), a recognized expert in teacher effectiveness, developed a framework for evaluating teacher

performance known as the Danielson Framework. Danielson has worked for 25 years as a teacher, curriculum director, staff development director, and assessment designer (Danielson, 2007). Her framework for teaching is currently utilized across the country.

The Danielson Framework (2007) was another of the approved evaluation tools and was the most widely chosen in the state. This framework includes 22 components categorized in four domains: planning and preparation, classroom environment, instruction, and professional responsibilities. The first domain, planning and preparation, and the third domain, instruction, each include six components. The second domain, classroom environment, and the fourth domain, professional responsibilities, each include five components. A teacher receives a rating of below basic to distinguished in each of the components which are then combined for an overall rating in each domain. Each teacher receives a score for each observation, rather than only a summative score at the end of the year. These observation scores are averaged for the summative evaluation and include the SGO(s) and/or the SGP.

The Marzano Teacher Evaluation Model. The Marzano Teacher Evaluation Model (2013), also approved by the DOE, includes four domains: classroom strategies and behaviors, planning and preparing, reflecting on teaching, and collegiality and professionalism. The first domain, classroom strategies and behaviors, includes 41 elements on which a teacher is rated. The second domain, planning and preparation, includes eight, and the third domain, reflecting on teaching, includes five. The final domain, professionalism, includes a total of six elements. A teacher can receive a rating on each element, and then through a combination of the elements, teachers are rated in each domain. Each teacher receives a score for each observation, rather than only a

summative score at the end of the year. These observation scores are averaged for the summative evaluation. The ratings system used by the Marzano Teacher Evaluation Model are beginning, developing, applying, or innovating (2013).

In 2014, Chuong wrote a critique about the inconsistency in the implementation of teacher evaluation and teacher reform across the country. While her data is collected in other states, she also brings up several important factors that can be applied to the reform efforts in New Jersey, including the advantages some teachers have over others and the differences in the implementation process from district to district. AchieveNJ (2013) and the approved evaluation tools were an attempt to make a previously subjective system objective and based on evidence.

City Public Schools and the Stronge TEPES

According to AchieveNJ (2013), each district is required to create a committee known as the District Evaluation Advisory Committee (DEAC). The main purpose DEAC is to oversee and guide the planning and implementation of the district's evaluation policies and procedures for both teachers and principals (AchieveNJ, 2013). Members of the DEAC included the Superintendent, a Special Education administrator, a parent, a Board of Education member, one or more central office administrators overseeing the evaluation process, one or more administrators conducting evaluations, and a representative of teachers at each level: primary and secondary. After these committees were formed, they were given the charge to choose an evaluation tool that would meet the specific needs of the district (AchieveNJ, 2013).

Mooney (2012) presents a predicament that many districts were in before AchieveNJ (2013) was fully implemented. The dilemma that districts faced was the

ability to separate the true quality and value of the evaluation tool versus the propaganda presented by vendors. As districts shopped for an evaluation tool, comparing evaluation systems such as The Danielson Framework (2007), The Marzano Teacher Evaluation Model (2013), and The Stronge TEPES (2006), superintendents were often met with sales pitches from the originators of the tools, as well as the companies creating technology to support the systems. However, the final decision of which evaluation tool would be implemented in the district came from the DEAC through a comparison of each tool in regards to the district's needs.

After forming a DEAC, City Public Schools, a pseudonym given to an urban school community in New Jersey, chose to implement the Stronge TEPES (2006), an NJ DOE approved evaluation tool, to manage the mandates of AchieveNJ (2013). In addition to City Public Schools, there are currently over 125 districts in the state of New Jersey who are using the Stronge TEPES (2006). This evaluation tool includes seven standards: professional knowledge, instructional planning, instructional delivery, assessment of/for learning, learning environment, professionalism, and student achievement. Using the Stronge Teacher Effectiveness Performance Evaluation System Handbook (2015) as a guide, evaluators throughout the state provide teachers with two or three observations, based on their tenure status, and one summative evaluation. Different from the other evaluation tools, the Stronge TEPES (2006) does not rate a teacher until the summative evaluation. At that time, a building level administrator, most often the principal, reviews all the evidence collected for each of the performance standards and applies a rating.

Each of the first six standards are weighted the same, while the seventh standard, student achievement, is weighted based on SGOs and/or SGPs. This allows evaluators to use a subjective point of view and voice in the final evaluation because they are able to wait until the summative evaluation to give a rating, which contrasts with the other evaluation tools that require ratings after each observation allowing for subjectivity and a lack of inter-rater reliability in the use of the Stronge TEPEs (2006).

Statement of the Problem

The criteria listed in the application for the federal Race to the Top grant (2009) have caused the purposes and goals of public education to come under scrutiny. This federal grant put a focus on reform in the following areas: the adoption of standards and assessments to prepare students to be college and career ready, the creation of a data system that measured student growth and success, the implementation of a system to recruit, develop, reward, and/or retain effective teachers and principals, and the execution of a plan to turn around the lowest performing schools (Race to the Top, 2009).

It is evident through these many reforms that politicians, educators, community members, parents, and students saw a need for change to the system that existed. The changes have affected both the culture and the dynamics of individual states and school districts. One of the most drastic of these changes, especially for teachers, came in the form of new evaluation systems, perhaps because of the lack of buy-in from both teachers and administrators (NJPSA, NJASA, & NJEA, 2014).

Despite tenure reform efforts and new teacher evaluation systems, it is my belief that teacher practice has not been positively impacted by the implementation of new evaluation systems. More than 50% of practitioners, including superintendents,

administrators, and teachers, all displayed concerns with the ability for AchieveNJ (2013) to accomplish its intended goals of improved quality instruction and increased student achievement (NJPSA, NJASA, & NJEA, 2014). Potentially, if we refocus our efforts in the area of teacher evaluation from only using it to comply with TeachNJ (2012) and AchieveNJ (2013) to help teachers grow professionally as Zepeda (2007) discussed in her instructional supervision framework, the Stronge TEPES (2006), as well as the other evaluation systems will likely have the potential to create positive change in education for administrators, teachers, and most importantly students, resulting in improved instruction and increased student achievement.

Purpose of the Study

The purpose of this mixed methods research study was to gather and analyze data from teachers in City Public Schools who have been observed and evaluated using the Stronge TEPES (2006). Through this mixed methods study, I investigated teacher perceptions of whether the Stronge TEPES (2006) was successfully implemented in the City Public Schools and if teachers felt they were receiving quality feedback that helped guide and improve instruction. The responses of teachers from this study will be used to inform future practice of administrators, as well as assist in the guidance of professional learning for teachers.

By using mixed methods approach, the qualitative and the quantitative data were used to provide information in order to draw a final conclusion about the effectiveness of the Stronge TEPES as perceived by educators. Using a survey that included questions aligned to a Likert scale, data were collected regarding the implementation process, the potential for the process to create positive change, the quality of feedback, and future

needs from the perspective of teachers in order to inform future practice and to improve administrator use of the Stronge TEPEs (2006). Follow up interviews, as well as an analysis of summative evaluation scores from year to year, were used to supplement the data, findings, and conclusions. The conclusions from this study may be used to drive professional growth and learning resulting from the evaluation results and teacher need. The following research questions guided my study:

1. How do teachers in the City Public Schools perceive the implementation, specifically the ability to create positive change, of the Stronge Teacher Effectiveness Performance Evaluation System?
2. How do teachers in the City Public Schools perceive the quality of feedback they are receiving in observations and evaluations using the Stronge Teacher Effectiveness Performance Evaluation System?
3. How do the summative evaluation scores of teachers in the City Public Schools demonstrate consistency in the use of the Stronge Teacher Effectiveness Performance Evaluation System?
4. How can the Stronge Teacher Effectiveness Performance Evaluation System be used to shift the supervisory model in City Public Schools from evaluation to instructional supervision and professional development?

Significance of the Study

This research is critically important based on the current state of education in New Jersey. The morale of teachers has declined, based either on the perception that teacher evaluation is punitive, or the lack of consistency and value in observations (NJPSA, NJASA, & NJEA, 2013). However, it is essential that we redirect our work in

the area of instructional supervision, using the framework that Zepeda (2007) outlines. By identifying the perceptions that currently exist, we, as educators, can reevaluate our practice and improve our work. Instead of completing classroom observations and summative evaluations as a way to be compliant with the regulations of TeachNJ (2012) and the mandates of AchieveNJ (2013), administrators, observers, and teachers need to use observations and summative evaluations as an avenue for professional growth and improvement of teachers, and, in turn, increase student achievement through improved instruction.

Practice

This mixed methods study has the potential to influence both teacher and administrator practice in the future. Teacher evaluation, as a result of AchieveNJ (2013), has created a major shift in the responsibilities of administrators. It has increased the number of times teachers are formally observed, and in turn, it has also increased the amount of observations that administrators are responsible for completing, decreasing the amount of time spent in other areas. When surveying superintendents, administrators, and teachers, NJPSA, NJASA, and NJEA (2013) found that each group had concerns about the administrators' capacity to accomplish what is outlined in TeachNJ (2012) and AchieveNJ (2013). To balance this situation, this study hopes to create a positive change in the use of teacher observations and evaluations. Rather than completing as a means for compliance, administrators can shift the focus of supervision of teachers to professional development and instructional supervision (Zepeda, 2007) that will result in both improved instruction and increased student achievement, two of the primary goals of AchieveNJ (2013).

Policy

TeachNJ (2012), as well as the accompanying regulations of AchieveNJ (2013), could be affected by the outcome of this mixed methods study, as well as the others completed based on these changes to practice and policy. As the NJ DOE continues to examine the policy, changes are being made. For example, when AchieveNJ (2013) was implemented in September 2013, tenured teachers were required to be observed a minimum of three times during each school year. However, in September 2016, the NJDOE changed the regulations to only mandate two observations of tenured teachers during a school year based on the feedback of the state's educators and administrators. Additionally, in a 2018 NJ DOE memo, the weighted components of a teacher's summative evaluation score were changed. At this time, according to AchieveNJ (2013), 5% of a teacher's summative evaluations score comes directly from a NJDOE created SGP, rather than the initial 30% proposed.

It is possible that this study can also create a change to the mandates, perhaps to include a stronger focus on the professional growth of teachers based on both observations and summative evaluations. It would be prudent for AchieveNJ (2013) to once again be revised to require a direct correlation between a teacher's summative evaluation and the following year's individual professional development plan. This would shift supervision of teachers along the continuum that Zepeda (2007) discusses from evaluation to instructional supervision and professional development.

Theoretical Framework

Minnear-Peplinski (2009) created a study with the goal to understand the process of supervising teachers. Using a mixed methods approach, Minnear-Peplinski (2009)

conducted a survey with questions aligned to the Likert scale, as well as open-ended questions, and collected data from participants throughout the country. Minnear-Peplinski (2009) attempted to survey all of the NASSP and NEASP Principals of the Year. In addition, for each principal surveyed, three teachers who this principal evaluated were surveyed and included in the study. Using this approach, Minnear-Peplinski (2009) was able to triangulate the data. Her sample size included 56 principals and 137 teachers.

The purpose of the study was to determine which supervisory method was most commonly used. The study looked at the perspective of both the principals and the teachers. Zepeda's (2007) three pronged model was used as the framework; the three elements of the framework are of instructional supervision, professional development, and evaluation. Zepeda (2007) defines: instructional supervision as that which "aims to promote growth, development, interaction, fault-free problem solving, and a commitment to build capacity in teachers" (p. 29), professional development as the teacher's or supervisor's focus on the development of professional expertise using problem solving and inquiry, and evaluation as judging the quality of a teacher's performance. Minnear-Peplinski (2009) concluded that evaluation was the least common form of supervision being utilized by the participants, with instructional supervision being the most common form of supervision.

Zepeda (2007) identifies the three main aspects of the supervisory process as instructional supervision, professional development, and evaluation. Similarly, Chen and McCray (2012) discuss the whole teacher approach to teacher supervision inclusive of professional development. Similar to the looking at the whole student, the whole teacher

approach looks to support the teacher in multiple ways. Both Zepeda (2007) and Chen and McCray (2012) emphasize the importance of differentiating between evaluation and supervision of teachers. Supervision of teachers is more than just observing and evaluating them; it is focused on helping them to improve through professional development (Chen and McCray, 2012). Shifting the goal and purpose of the evaluation system to instructional supervision and professional development as Zepeda (2007) defines will allow administrators to focus their supervision of teachers on a more positive approach.

The purpose of teacher evaluation in the state of New Jersey, inclusive of both TeachNJ (2012) regulations and the AchieveNJ (2013) mandates, is to improve both teacher practice and to increase student achievement (AchieveNJ, 2013). With the focus of improving teacher practice and providing teachers with an opportunity to access quality professional development opportunities at the forefront of this study, teacher evaluation, including its ability to bring positive change, should focus on adult learning and teacher development. The feedback provided to teachers should give a direction toward lifelong learning and improvement in order for the focus to be on continual improvement and growth for all. Chen and McCray (2012) provide a theoretical framework that focuses on professional development and learning for teachers. A shift in culture is needed to move away from punitive measures based on teacher evaluation to learning and growth for educators. Our focus needs to change from an emphasis on catching educators doing the wrong thing to supporting them and providing them opportunity to succeed. Administrators may need to shift the type of feedback provided

in order to accomplish this, if improved instruction through teacher learning and growth is the goal.

Definition of Terms

The following terms are defined in terms of this study's purpose:

AchieveNJ. AchieveNJ (2013) is the improved educator evaluation and support system created to better align educator evaluation with practices that lead to improved student outcomes (AchieveNJ, 2013).

Documentation log. The documentation log includes both specific required artifacts and teacher selected artifacts that provide evidence of meeting selected performance standards included in the Stronge TEPES (Stronge & Associates, 2015).

Evaluation. Evaluation is defined as judging the quality of a teacher's performance. Evaluation is part of supervision (Zepeda, 2007)

Instructional supervision. Zepeda (2007) defined instructional supervision as that which "aims to promote growth, development, interaction, fault-free problem solving, and a commitment to build capacity in teachers (p. 29)".

Inter-rater reliability. Inter-rater reliability ensures that all observers are consistent in their ability to identify and rate observations with reliability and consistency among evaluators (Stronge & Associates, 2015).

NJ ASK. The New Jersey Assessment of Skills and Knowledge (NJ ASK) was the state assessment for students in grades 3 through 8. The NJ ASK measured student achievement in the areas of ELA and Mathematics. Additionally, students are assessed in the area of science in grades 4 and 8. The only remaining portion of NJ ASK given to

students in the state is the science assessment in grades 4 and 8 (New Jersey Department of Education [NJ DOE], 2014).

Non-tested teacher. A non-tested teacher is any teacher who does not fit the criteria outlined for a tested teacher during the school year of the evaluation (AchieveNJ, 2013).

Observation. Observations are an important source of performance information. Formal classroom observations focus directly on the Stronge TEPES (2006) teacher performance standards. Informal observations or walk-throughs are intended to provide more frequent information on a wider variety of contributions made by the teacher. Evaluators are encouraged to conduct observations by visiting classrooms, observing instruction, and observing work in non-classroom settings (Stronge & Associates, 2015).

PARCC. The Partnership for Assessment of Readiness for College and Careers (PARCC) is a consortium of states that collaboratively developed a common set of assessments to measure student achievement and preparedness for college and careers in the areas of ELA and Mathematics. The PARCC Assessments are aligned to the Common Core State Standards (CCSS) and were created to measure students' ability to apply their knowledge of concepts rather than memorizing facts (PARCC, 2016).

Professional development. Professional development is defined as the teacher's or supervisor's focus on the development of professional expertise using problem solving and inquiry (Zepeda, 2007).

Student Growth Objective. Student Growth Objectives, SGOs, are long-term academic goals for groups of students set by teachers in consultation with their

supervisors (AchieveNJ, 2013). These objectives or goals are based on the achievement levels of students on either benchmark assessments or a pre/post assessment.

Student Growth Percentile. Student Growth Percentiles, SGPs, are a measure of how much a student improves his/her state test performance from one year to the next compared to students across the state with a similar score history (AchieveNJ, 2013). Currently, a teacher's SGPs is based on the median of a minimum of twenty students' performance on PARCC.

Summative evaluation. Teacher evaluation consists of two primary components: teacher practice (measured primarily by classroom observations) and student achievement (measured by SGOs and SGPs). A teacher's summative evaluation is rated using a four-point scale (AchieveNJ, 2013).

TeachNJ. TeachNJ (2012) is the tenure reform signed into law. The goal of the law is to increase student achievement through improved instruction. TeachNJ (2012) changes the processes of earning and maintaining tenure by improving evaluations and opportunities for professional growth (New Jersey Department of Education [NJ DOE], 2014).

Tested teacher. A tested teacher is one who is assigned to a grade 4 through 8 ELA or Mathematics class for 60% or more of the school year prior to the date on which the state standardized test is administered. In addition, a tested teacher is assigned 20 students during the school year of the evaluation or the combination of up to two previous years plus the current year. These twenty students must be enrolled in the class for 70% of the year prior to the administration of the state standardized test (AchieveNJ, 2013).

Conclusion

The Race to the Top grant (2009), as well as the amendment to TeachNJ (2012) and the mandates of AchieveNJ (2013), have changed the landscape of education and teacher evaluation in the state of New Jersey. Teachers, administrators, and superintendents in the state have seen shifts in their responsibilities and expectations because of each of these (NJPSA, NJASA, & NJEA, 2013). However, it is possible that the supervision of teachers has not been positively impacted. The primary goals of improving instruction and increasing student achievement (AchieveNJ, 2013) by adopting an evaluation system with a four point rating scale should be the focus of educator practice, rather than compliance. Like all other districts, City Public Schools implemented a teacher evaluation system, the Stronge Teacher Effectiveness Performance Evaluation System (2006), in September 2013. The purpose of this study was to examine teacher perceptions of the implementation, its ability to affect positive change in instruction and achievement, and the connection between evaluation and professional growth.

The following chapters review the body of literature that exists, the methodology utilized to complete the study, the findings of the study, and the conclusions of the study. The implications and recommendations for practice, policy, and future research will also be examined.

Chapter 2

Literature Review

The American Recovery and Reinvestment Act (ARRA), signed by President Barack Obama in February 2009, created a competitive grant known as Race to the Top (RTT). With 4.35 billion dollars available to individual states, RTT focused on four criteria: the adoption of standards and assessments to prepare students to be college and career ready, the creation of a data system that measured student growth and success, the implementation of a system to recruit, develop, reward, and/or retain effective teachers and principals, and the execution of a plan to turn around the lowest performing schools (Race to the Top, 2009). These criteria led to K-12 reform efforts throughout the country.

Under Governor Chris Christie's administration, New Jersey applied for and received RTT funding. The acceptance of this reward money led to the state: implementing and adopting the Common Core State Standards, joining the PARCC consortium, amending the TeachNJ (2012) regulations, mandating AchieveNJ (2013), and including a data system for student growth known as student growth percentiles (SGPs) in teacher evaluations.

As various states in the country continue to make changes to teacher evaluation, research and studies continue to attempt to determine the effects of teacher evaluation reform on both instruction and student achievement. According to Doherty and Jacobs (2015), in 2015, there were just five states, California, Iowa, Montana, Nebraska and Vermont, that still had no formal state policy requiring that teacher evaluations take

objective measures of student achievement into account in evaluating teacher effectiveness.

Completed studies, including Shepard (2013), Curran (2014), Winslow (2015), Towe (2012), and Wacha (2013), are looking at the teacher and administrator perceptions of the evaluation systems that are currently in place. Shepard (2013) studied the perceptions of teachers and administrators regarding the teacher evaluation process in the state of Georgia. Winslow (2015) looked at the perceptions of teachers and administrators regarding the feedback using a Danielson-based teacher evaluation system. Towe (2012) investigated the role of a teacher evaluation system and its influence on teacher practice and professional growth in four urban high schools in New Jersey. Wacha (2013) examined teacher perceptions of the entire teacher evaluation process following the amendment to TeachNJ (2012).

The purpose of this convergent, parallel mixed methods research study was to gather and analyze data from teachers who have been observed and evaluated using the Stronge Teacher Effectiveness Performance Evaluation System (TEPES) in City Public Schools, with regards to it being able to create positive change in both quality instruction and student achievement, as well as the quality of feedback received by teachers on observations and summative evaluations. While this study focuses primarily on the perceptions of the teachers in the district, future administrator practice may be influenced by the findings. This chapter will examine and synthesize the existing literature on the topical areas of teacher evaluation, the Stronge TEPES, and instructional supervision in order to be able to draw conclusions, understand the implications, and make recommendations.

History of Teacher Evaluations in the United States

In order to understand current teacher evaluation practices, a review of the history of teacher evaluations is necessary because past practices have a direct correlation to the path we are taking today. Marzano, Frontier, and Livingston (2011) outline the history of teacher supervision and evaluation. This historical perspective begins in the 1700s with members of the clergy guiding educators (Marzano, Frontier, & Livingston, 2011). During this time, supervisory committees were given the charge to monitor instruction, and because they had almost absolute control, the ways in which supervision occurred varied drastically from one place to another (Marzano et al., 2011).

During the 19th century, there was a shift in education from religious controlled schooling to state and government controlled schools (Marzano et al., 2011). At this time, the country saw the development of mass, or universal, education and common schools, with two trains of thought as to why this occurred (Marzano et al., 2011). The first belief is that common schools were a result of industrialization which would teach future workers a respect for law and authority, and the second belief is that common schools and mass education were a way to ensure the continuation of colonial Puritan enthusiasm so that everyone could read the Bible (Vinovskis, 1992).

The 1940s and 1950s saw a focus on teacher traits such as appearance, approachability, enthusiasm, truth worthiness, and emotional stability (Danielson & McGreal, 2000). When evaluating teachers, these were the criteria deemed most important. If a teacher exhibited these personal characteristics, they were effective in the eyes of the evaluator; none of these focused on the quality of instruction (Danielson & McGreal, 2000).

As school systems and their purposes became more complex, education in the country took on a form of scientific management until after World War II (Marzano et al., 2011). Scientific management, based on the work of Frederick Taylor, attempted to look at education through a similar lens that factory work was viewed (Marzano et al., 2011). For example, if there were several ways in which to accomplish something, Taylor argued that one was more effective (Bouie, 2012). This correlated to the belief that there was one right or more effective way to teach. This time period brought about the teacher leaders and an increase in the responsibilities of supervisors in order to identify the most effective form of teaching (Marzano et al., 2011). Teachers were hired based on specific skills sets and areas of expertise (Marzano et al., 2011).

Education in the country shifted again in the late 1960s, perhaps as a result to the changes in the economy and society. Doerr (2012) emphasizes that the skills considered to be the most valuable in a teacher during the 1960s and 1970s were math and science. By the 1980s and 1990s, Madeline Hunter's model became the most emphasized guide (Marzano et al., 2011). Hunter's model centers on the elements of instruction. According to Hunter (2004), these elements included the anticipatory set, the lesson's objective, input modeling, check for understanding, guided practice, independent practice, and closure (Hunter, 2004). Clinical supervision also became more common (Doerr, 2012) and gave a prescriptive approach to teacher supervision, which then shifted to a more reflective approach based on a study conducted by the RAND group. This study, titled *Teacher Evaluation: A Study of Effective Practices*, attempted to determine what practices were occurring at that time, as well as identifying problems that evaluators and teachers saw with the system (Marzano et al., 2011).

A drastic change to the evaluation process came after the National Commission of Excellence in Education's 1983 report *A Nation-At Risk: The Imperative for Educational Reform* (Gardner, 1983). This report asserted that public education in the nation was in dire need of reform. Because of this, teacher evaluation began to be based on judging classroom performance against a set of criteria or behaviors (Medley & Coker, 1987). During this time, Charlotte Danielson introduced the Danielson Framework (Danielson, 2007). The Danielson Framework (2007) began as a guide for teaching responsibilities grounded in the constructivist view of learning and teaching. While still a framework for teaching, Danielson has made adjustments to the original framework so it could now be used as an evaluation tool, rather than just a teaching framework. This is because since the beginning of the 21st century we have seen the emphasis change from teacher supervision to teacher evaluation (Marzano et al., 2011).

As the country focuses on education reform in the areas of both tenure and compensation, Rucinski and Diersing (2014) portray the role of RTT (2009) in the history of evaluation through advances such as value-added modeling, which is a statistical model to compare actual student growth with expected student growth. Through a look at the addition of a value-added modeling to teacher evaluation, they explain that RTT (2009) attempts to tie budgetary increases to the implementation of objective evaluation systems and performance pay (Rucinski & Diersing, 2014). If ultimately teacher pay and retention is affected by evaluation systems, there needs to be a way to ensure accuracy.

History of Teacher Evaluations in New Jersey

For more than three decades before the availability of the Race to the Top grant (2009), the amendment to TeachNJ (2012), or the regulations of AchieveNJ (2013),

teacher evaluations throughout the state were based on minimum state mandates. TeachNJ (2012) is the law regarding tenure, and AchieveNJ (2013) describes the regulations detailing the supports structures needed to follow TeachNJ (2012). Because each district has its own Board of Education as its governing body, each adopted its own unique evaluation tool. During the election, Governor Chris Christie ran on the platform that education in New Jersey, as well as the country, was stagnant (Rundquist, 2009), which became the driving force beyond tenure reform.

Prior to the 2012 amendment to TeachNJ (2012), tenured teachers only needed to be observed once annually, and non-tenured teachers had to be observed three times throughout the course of the school year, with a minimum occurring once during the first semester and once during the second semester (TeachNJ, 2012). These observations, in addition to the subjective perspective of administrators, were used to complete an end of the year evaluation. However, there were no specific requirements about what needed to be included in an observation or an evaluation.

Much of the history of teacher evaluation in the state is based on the history of tenure in New Jersey. New Jersey was the first state to pass tenure legislation in 1910, which mirrored the labor movement in the country (Stephey, 2008). While the first legislation applied to college professors, the women's suffrage movement created the tenure movement in K-12 education, in an attempt to stop female teachers from losing their jobs based on pregnancy (Stephey, 2008). While tenure provided teachers with protection from unjust firing, it also brought about the difficulty of removing a tenured teacher from his/her position even for just causes (Stephey, 2008). The cost of removing a tenured teacher became so large based with legal fees that the evaluation of teachers

became almost irrelevant (Stephey, 2008). In the Fleisher (2010) article, Governor Christie stated, “Tenure has become a job guarantee regardless of performance or success. Tenure has become the sclerosis that coats the veins of our school system.”

The amendment to TeachNJ (2012) and the adoption of AchieveNJ (2013) were derived from national reform efforts in K-12 public education, beginning with the No Child Left Behind of 2002 and the Race to the Top competitive grant in 2009, all of which occurred during Governor Christie’s terms in office.

Reform Efforts

The federal and state governments played a role in the reforms efforts and changes that occurred in public education. Through the adoption and revision of federal and state policies, the requirements of local education agencies was affected. In recent history, No Child Left Behind (2001) and Race to the Top (2009) generated the greatest reforms efforts, including college and career ready standards, college and career ready assessments, and the implementation and adoption of value added teacher evaluation systems.

No Child Left Behind. In 2001, the Elementary and Secondary Education Act (ESEA) was reauthorized through the No Child Left Behind Act (NCLB). This act allowed the federal government to continue to have a role in the education at the district level. Through NCLB (2001), there was increased accountability in public education, support for standards based reform, a stronger emphasis on annual standardized testing in schools, the measure of academic progress, and the monitoring of teacher quality (Superfine, Gottlieb & Smylie, 2012). While the Act did not explicitly describe the ways in which individual states had to accomplish these goals, NCLB (2001) did attach federal

funding to encourage compliance with them (Manna, 2006). NCLB (2001) was once again reauthorized in 2010 by President Barack Obama. In this reauthorization the criteria of the competitive Race to the Top grant (2009) were included.

Race to the Top. On February 17, 2009, the American Recovery and Reinvestment Act (ARRA) was signed into law by President Barack Obama (Race to the Top, 2009) because of the financial crisis the country experienced (McQuinn, 2012). The ARRA created a competitive grant totaling 4.35 billion dollars known as Race to the Top (RTT). Through an application process, the grant was made available to individual qualifying states. The funding was meant to create innovation and reform in local K-12 districts in each state. Each state was able to submit an application and was awarded points based on their inclusion of four criteria: the adoption of standards and assessments to prepare students to be college and career ready, the creation of a data system that measured student growth and success, the implementation of a system to recruit, develop, reward, and/or retain effective teachers and principals, and the execution of a plan to turn around the lowest performing schools (Race to the Top, 2009). Race to the Top (2009) led to the state of New Jersey, as well as other states in the country, adopting the Common Core State Standards and becoming part of the PARCC Consortium.

Common Core State Standards. In 2010, New Jersey, along with 41 other states, adopted the Common Core State Standards (CCSS) (Common Core State Standards Initiative website, 2014). These standards were a direct result of the criteria outlined in the Race to the Top grant (2009). The CCSS in mathematics and English Language Arts set forth what students need to have learned by the end of each year and how teachers need to update, alter, and amplify their own knowledge base and

pedagogical arsenal if they are to succeed in imparting the CCSS to their students (Porter, McMaken, Hwang, & Yang, 2011).

Kurabinski (2010) details three instructional shifts that resulted from the CCSS in English Language Arts. These include regular practice with complex text, reading, writing, and speaking grounded in evidence from texts (both literary and informational), and building knowledge through content-rich nonfiction. Educators have designed curriculum and chosen resources with these shifts in mind. These instructional shifts have been furthered reiterated through lesson plans and unit plans available to all educators (Achieve the Core, n.d.). Additionally, educators have access to a common set of lessons that can be implemented in any classroom around the country.

Porter, McMaken, Hwang, and Yang (2011) state, “the Common Core Standards released in 2010 represent an unprecedented shift away from disparate content guidelines across individual states in the areas of English language arts and mathematics” (pg. 103). Despite the push for commonality, districts have the ability to design individual curricula based on the standards.

PARCC. Race to the Top (2009) also included a criteria to ensure that students were being assessed to determine college and career readiness. Achievement is measured in part using The Partnership for Assessment of Readiness for College and Careers (PARCC), standardized assessment aligned to the CCSS (PARCC, 2014). This assessment differed drastically from the basic skills assessments students were used to taking. This assessment required students to spend hours on a device writing synthesis essays and completing technology enhanced mathematical problems (PARCC, 2014).

In 2018, New Jersey once again shifted the way in which student achievement is measured and now is mandating that students in grades 3 to 11 take the New Jersey Student Learning Assessment in the spring of each school year (Spring Assessment, 2018). This assessment looks much like the original PARCC assessment; however, it is a New Jersey specific assessment.

Teacher evaluation reform. In addition to new standards and a new assessment system, public education in New Jersey saw reform in the area of teacher evaluation because of the acceptance of Race to the Top (2009) funding. In August 2012, New Jersey amended the TeachNJ Act, a law regarding tenure for teachers, vice/assistant principals, and principals in the state. TeachNJ (2012) increased the number of years of effective service in order for tenure to be awarded to four years and one day. Summative evaluation scores for teachers, principals, and assistant principals must be either effective or highly effective in three of these years (TeachNJ, 2012)

Public school districts are also required to adhere to the mandate of AchieveNJ (2013), that details and support structures necessary to carry out the TeachNJ (2012) law. Each district was required to adopt and implement a teacher evaluation model that included both teacher practice and student achievement. Each of these three evaluation systems included a four-point rating scale, which was one of the primary requirements of AchieveNJ (2013).

Teacher Point of View on Evaluation

The purpose of this research study was to examine the perceptions of teachers who are observed using the Stronge TEPEs (2006) implementation process, with regards to being able to create positive change in both instruction and student achievement, as

well as teacher perceptions of the quality of feedback they are receiving on observations and evaluations. Several other studies of a similar methodology have been conducted and focused on other teacher evaluation tools. As it was important to look at the history of teacher evaluation in the country, it was also important to synthesis the literature and studies that exist on teacher evaluation throughout the country.

Shepard (2013) used a mixed methods approach to collect and analyze qualitative and quantitative data from 277 teachers and 12 administrators in the state of Georgia. Participants were from three small, rural school systems and were surveyed using both Likert scale and one open-ended question based on a new teacher evaluation tool in Georgia (Shepard, 2013). Shepard (2013) concluded that teachers were reasonably satisfied with the evaluation tool and the observations they were receiving; 91% of the participants in the study felt that the overall quality of teacher evaluation was average or above average quality. Specifically, teachers felt that the most significant component of the evaluation system was the feedback they were receiving from administrators (Shepard, 2013). Additionally, teachers expressed that the evaluations were based on individual and institutional improvement that could better serve the students and the community (Shepard, 2013).

Further, Curran (2014) used a mixed methods approach and examined the perceptions of teachers in Texas, using the Professional Development & Appraisal System (PDAS). Curran's sample included 94 teachers from 13 elementary schools. Only elementary teachers were included in the sample. In addition to using a survey, Curran (2014) also interviewed six teachers and created an online focus group. The conclusions from this study focused on the time-constraints that existed because of

PDAS. Teachers expressed that the time constraints led to less formative assessments (of their instruction), as well as less in-depth feedback. Curran stated, “the three main features of productive feedback are as follows: high specificity, high frequency, and most importantly, high positivity. The participants stated on several occasions that they perceive a need for feedback data to be very specific. In order for instructional changes to be made, teachers need to know exactly what they do well and need to continue to do in their classroom, what practices are evolving and need to be fostered, and what things are unconstructive and not beneficial to student achievement. (pg. 108)”. Similar to the findings of Shepard, teachers perceive high quality feedback as necessary for evaluations to have an impact on instruction.

Winslow (2015) conducted a mixed methods, action research study about the perceptions of the implementation of the Danielson Framework (2007) in Illinois. The Danielson Framework (2007) was one of the initial options presented to districts in the state of New Jersey. Participants in Winslow’s (2015) study included teachers, principals, and vice principals from one school district. Using data collected from interviews and an online survey, Winslow concluded that teachers saw the feedback provided by administrators as different from what they had received in the past. Additionally, teachers expressed that the feedback was better because it was formative. More specifically, Winslow found that the “feedback received following an observation was different and better in the new evaluation system compared to prior methods experienced by teachers. (pg. 40)”. Participants indicated that by using a standardized evaluation tool the feedback is more frequent, more prompt, and more objective. Also notable was that teachers expressed concern about the process of the observations and

evaluations and that they had concerns about the amount of training they had received prior to implementation (Winslow, 2015).

While these studies occurred outside of the state of New Jersey and were not based on the usage of the Stronge TEPES (2006) in a district, there are notable similarities between the circumstances surrounding teacher evaluations in these cases and what has happened in the past several years in City Public Schools. One similarity between these three studies and this study was the way in which data was collected. In each, data was collected using a survey. Another similarity was that the quality of feedback teachers are receiving appears to be at the forefront of the conversations and data.

It is also essential to view this study in the context of other studies that were completed in the state of New Jersey, especially based on the teacher evaluation reform efforts that have occurred in the last decade under the tenure of Governor Chris Christie. Prior to the required adoption and implantation of a research based teacher evaluation system in September 2013, districts in the state had the opportunity to pilot one of the recommended tools. Ten districts completed a pilot in 2011 which further informed policy decisions in the state (Wacha, 2013).

Towe (2012) conducted a mixed methods research study in four urban high schools who participated in the pilot of the Danielson Framework (2007). From these four high schools, 30 participants completed a questionnaire and were interviewed. In addition, Towe (2012) included NJ DOE documents in her study. Towe's (2012) data suggests that teachers saw the purpose of teacher evaluation as two-fold: a way to increase accountability and a way to foster growth in teachers. She determined that

teachers perceived the observation and evaluation system as having limited impact on their ability to improve classroom practices. However, teachers did see the summative evaluations as having the potential to bring about quality professional development. Teachers also expressed that they have a strong level of trust in the feedback they received from administrators. Towe (2012) stated, “Data suggest that teachers perceive that the teacher evaluation process has some degree of impact on their teaching practice, while they perceive the process as having little impact on their professional growth. (pg. 147)”. Despite that the participants saw one of the purposes of teacher evaluation as a way to foster growth, their perception is that this is not actually occurring. Further proofing Towe’s (2012) conclusion that the perception of teachers is that the teacher evaluation process as perfunctory and done simply to satisfy district and state mandates.

Wacha (2013) used a mixed methods, action research approach to determine the perceptions of ten teachers from one high school in New Jersey. Data were collected from participation in a survey, an open-ended questionnaire, and semi-structured interviews. In this research study, the overall perceptions of teachers had a negative connotation, including that they saw the evaluation systems as not an effective measure of teachers’ performance. The goal of teacher evaluation is to improve teaching (Taylor & Tyler, 2012). Wacha (2013) concluded that teachers would only support teacher evaluation reform if they perceived the true goal as a way to improve instruction and student achievement rather than a way to punish teachers or hold them accountable. Wacha (2013) concluded that this could be contributed to overall climate in the state during this shift, potentially created by the NJEA opposition to the policies of AchieveNJ (2013). Governor Christie criticized the NJEA calling the union “a group of political

thugs” (Blackburn, 2011). Wacha’s (2013) final conclusions were that teachers received minimal feedback to guide improvement of instruction and that there was no direct correlation between the feedback they received and future professional development.

These two studies, while taking place in New Jersey, do not indicate teacher perceptions of the Stronge TEPES (2006). However, they both addressed the lack of follow through from either evaluators or districts after observations or summative evaluations. The studies showed a lack of correlation between teacher evaluation and professional development opportunities, resulting in minimal growth in the area of improved instruction.

Administrator Point of View on Evaluation

Another purpose of this research study is to improve the future administrator use of the Stronge TEPES (2006) implementation process from compliance to one of informed effective practice able to create positive shifts in both teacher instruction, teacher professional development, and student achievement.

Shepard (2013) found that, in contrast to her original assumption, administrators did not have an overwhelming negative perception of the teacher evaluation system. Of the 12 administrators surveyed in the study, nine of them were from the same district, which may have skewed the results and conclusions. Only one administrator included in the study, not from the same district, found the evaluation process to be ineffective. Shepard (2013) found that administrators perceived teacher evaluations to have the least impact on school climate and professional growth, while they perceived teacher evaluations to have the most impact on student learning. While Shepard (2013) found that administrators thought the timing of feedback received by teachers was the greatest

attribute of the implementation of an evaluation system, they only perceived the feedback provided as adequate.

In the study completed by Curran (2014), the majority of the data came from teachers; however, administrators were also included in the online focus group. In this format, administrators expressed similar concerns as the teachers, less in depth feedback being provided on both observations and evaluations. Because of the time constraints that administrators were dealing with, they felt that they were not able to provide teachers with individualized feedback. Curran (2014) concluded that the lack of time and the lack of individualized feedback were limiting the professional growth for teachers based on observations. Curran found that administrators perceived, “formative evaluation and feedback as a necessity for professional growth. (pg. 102). The administrators in the study felt that the feedback they were providing was infrequent, non-specific, and at times, overly negative. With a willingness to change and the understanding that change was a necessity to improve instruction, Curan (2014) concluded that current, high frequency, and high quality, feedback must be provided to teachers.

The administrators included in Winslow’s (2015) study were those in direct contact with the teachers who participated. All of the administrators, principals and vice principals, were from the same district allowing for triangulation. Through Winslow’s (2015) survey, administrators noted that the most valuable feedback they were providing came in the form of conversations following the observations, rather than the actual written report. Winslow (2015) concluded that administrators perceived teachers as needing to reflect on their practice in order for teacher evaluation to have an impact on instruction. The perception of all of the administrators in Winslow’s (2015) study was

that the administrative time needed to implement the new evaluation system is immense. Administrators spoke directly to the scheduling concerns that exist because of the implementation of a teacher evaluation system. If administrators are to be instructional leaders (Shields, 2010), they must use observations and evaluations as a way to provide constructive feedback rather than merely a means to comply with mandates and regulations.

Towe's (2012) findings led to several conclusions regarding administrator perceptions of the evaluation system, specifically the Danielson Framework (2007) in this study. The administrators included as participants in this study were those from four urban high schools. Towe (2012) concluded that the administrators saw the evaluation tool as a means to provide teachers with directed professional development based on their specific needs. She also found that administrators saw the process as having a limited impact on improved practice, similar to the perceptions of the teachers in the study. Each administrator who participated in the study spoke to being responsible for completing observations and evaluations, as is mandated by state regulations and now district policy. However, only one administrator mentioned her role in improving instruction when discussing the evaluation system. Towe (2012) reported her saying, "I see my role as coach, colleague ... [and] instructional leader. (pg. 123)". Though administrators expressed confidence in their knowledge and capabilities to adequately and positively evaluate teachers, the lack of comments regarding instructional leadership led Towe (2012) to different conclusions.

The NJ DOE requested that the Rutgers University Graduate School of Education complete an assessment of the pilot program in the state. In August 2011, ten districts

were notified of their participation in the pilot. Beginning in January 2012, teachers in each of these districts were observed and evaluated using one of the NJ DOE approved tools. Firestone, Blitz, Gitomer, Kirova, Shcherbakovi, and Nordon (2013) began their study in March 2012 by surveying both teachers and administrators in those districts. A total of 2495 teachers were surveyed.

In their findings, Firestone et al. (2013) found that administrators looked for evaluation rubrics that were accurate, fair, provided useful feedback to help teachers improve their practice, and that could be used for personnel decisions. Administrators in the pilot had a positive view of the rubrics used in the teacher evaluation system; 74% of administrators agreed that the evaluation rubrics assessed teachers accurately. Similarly, 75% of administrators agreed that the rubrics generated information that provided useful individual feedback or guidance for professional development.

Despite that teachers and administrators could agree on what should be included in an evaluation rubric, the ways in which the two groups applied these criteria were drastically different according to Firestone et al. (2013). Administrators perceived teacher evaluation systems as providing teachers with accurate and fair assessments, as well as useful feedback to improve instructional practices.

Supervision of Teachers

In Minnear-Peplinski's (2009) study regarding the process of supervising teachers, she used a survey with questions aligned to the Likert scale, as well as open-ended questions, in order to collect data from teachers throughout the county. The purpose of the study was to determine which supervisory method was most commonly used by looking at the perspective of both administrators and teachers.

Zepeda's (2007) three pronged model was used as the framework for both Minnear- Peplinski's (2009) study and this study; Zepeda (2007) identifies the three main aspects of the supervisory process as instructional supervision, professional development, and evaluation. Zepeda (2007) defines: instructional supervision as that which "aims to promote growth, development, interaction, fault-free problem solving, and a commitment to build capacity in teachers" (p. 29), professional development as the teacher's or supervisor's focus on the development of professional expertise using problem solving and inquiry, and evaluation as judging the quality of a teacher's performance.

Chen and McCray (2012) discuss the whole teacher approach to teacher supervision which is inclusive of professional development opportunities. Both Zepeda (2007) and Chen and McCray (2012) highlight the significance of differentiating between teacher evaluation and the supervision of teachers. Supervision of teachers is more than just observing and evaluating them; it is focused on helping them to improve through professional development opportunities (Chen and McCray, 2012). Shifting the goal and purpose of the evaluation system to a focus on both instructional supervision and professional development will allow administrators to focus their supervision of teachers on a more positive approach.

The purpose of teacher evaluation in the state of New Jersey, inclusive of both TeachNJ (2012) and the AchieveNJ (2013) mandates, is to improve both teacher practice and student achievement (AchieveNJ, 2013). With the focus of improving teacher practice and providing teachers with an opportunity to access quality professional development at the forefront of this study, teacher evaluation, including its ability to bring positive change, should focus on adult learning and teacher development. The

feedback provided to teachers should give a direction toward lifelong learning and improvement in order for the focus to be on continual improvement and growth for all. As Robinson (2009) suggested, supervision is only successful when the shift is made from inspection to helping the teacher. Despite the intended goals of reform efforts, the practice has been on inspection.

In order to span all three levels of Zepeda's (2007) framework, administrators need to evaluate teachers, provide instructional supervision, and guide the professional growth and development of teachers. When teachers are informed about the evaluation process and the feedback they receive leads to professional growth, improved instruction, or increased student achievement, observations and evaluations are a map and path to professional development and improved instructional supervision; two aspects of the supervisory model that Zepeda (2007) discusses.

Currently, the usage of the Stronge TEPES (2006), in the City Public Schools, can be considered evaluation, the least common form of supervision found by Minnear-Peplinski (2009). However, if the district shifts the usage of the Stronge TEPES (2006) to instructional supervision and professional development both perception and culture could change. Teachers may feel more open to the feedback being provided in observations and summative evaluations if they perceive them as beneficial rather than punitive. Papay (2012) argues that an effective evaluation system must include a component that addresses ongoing teacher development and improvement, and that "assessment tools must be able to measure accurately and be unbiased, valid and reliable" (p. 127). They may also be open to evidence that creates a direct correlation between observations and evaluations and personalized professional growth.

Conclusion

This chapter presented a synthesis of the existing literature in the areas of both teacher supervision and teacher evaluation. It was essential to have an in-depth understanding of the history of teacher evaluation in both the country and the state before being able to understand the impacts of each of the recent reform efforts that have taken place across K-12 public education.

The perceptions and points of views of both teachers and administrators have been examined through various mixed methods and action research studies. The quality of feedback provided to teachers and the immense amount of time spent by administrators is an overarching finding in each of these studies. Through the successful implementation, adoption, and usage of a teacher evaluation tool, districts will begin to see supervision of teachers span each of the levels described in Zepeda's (2007) framework, each of which plays an integral role in the supervision of teachers.

Chapter 3

Methodology

In 2009, President Barack Obama, after signing the American Recovery and Reinvestment Act, dedicated 4.35 billion dollars to the Race to the Top (RTT) (2009) grant. This competitive grant was available to states that were willing to make reform efforts aligned to the primary purposes outlined in the application. The four criteria were: the adoption of standards and assessments to prepare students to be college and career ready, the creation of a data system that measured student growth and success, the implementation of a system to recruit, develop, reward, and/or retain effective teachers and principals, and the execution of a plan to turn around the lowest performing schools (Race to the Top, 2009).

As a way to ensure this funding, the state of New Jersey made an amendment to TeachNJ (2012) which included the mandates of AchieveNJ (2013). Beginning in September 2013, all districts were required to choose, adopt, and implement an evaluation model which included a four-point rating scale for a teacher's final evaluation (AchieveNJ, 2013). It also included student achievement in each teacher's summative evaluation through the inclusion of both student growth objectives, (SGOs), and student growth percentiles (SGPs) (AchieveNJ, 2013).

In September 2013, City Public Schools, a pseudonym given to maintain confidentiality, adopted and implemented the Stronge Teacher Effectiveness Performance Evaluation System (TEPES) (2006). The Stronge TEPES (2006), one of the NJ DOE approved systems, includes seven standards: professional knowledge, instructional planning, instructional delivery, assessment of/for learning, learning environment,

professionalism, and student achievement. The four-point rating scale for this evaluation tool includes: ineffective, partially effective, effective, and highly effective (Stronge, 2006).

The purpose of this mixed methods study was to examine teachers' perceptions of the Stronge TEPEs (2006) with regard to its ability to affect positive change in instruction, student achievement, as well as provide professional growth opportunities for teachers based on individualized feedback in observations and on summative evaluations.

Despite tenure reform efforts and new teacher evaluation systems through TeachNJ (2012) and AchieveNJ (2013), teacher practice has not been positively impacted by the implementation of new evaluation systems (NJPSA, NJASA, NJEA, 2013).

Through this mixed methods study, I investigated teacher perceptions of whether the Stronge TEPEs (2006) was successfully implemented in the City Public Schools and if teachers felt they were receiving quality feedback that helped to guide and to improve instruction. The responses of teachers may be used to inform future practice of administrators, as well as assist in the guidance of professional learning for teachers. In addition, it may change the supervision of teachers from merely being evaluative to a focus on instructional supervision and professional development, as Zepeda (2007) suggests.

The measures being used to observe and evaluate teachers remain objective and administrators have continued to rate teachers effective for the purpose of compliance rather than change. Because of this, it was necessary to develop an evaluation instrument that can gather data about teachers' perceptions of the implementation process, including its ability to create positive change through the quality of the feedback they received. In

order to gather honest data from teachers, the survey took close look at the evaluation system and its limitations for both teachers and administrators.

In accordance with AchieveNJ (2013), City Public Schools, as well as many other districts, adopted the Stronge TEPES (2006), one of the available approved tools, in September 2013. The evaluation system includes seven performance standards; the first six standards relate directly to teacher practice while the seventh is based on both SGOs and SGPs, measures of student achievement. The primary goal of the evaluation system is to provide teachers with objective observations and an objective summative evaluation each year based on a preponderance of evidence, in the form of observations, documentation logs, student surveys, SGOs, and SGPs for the purpose of improved instruction and professional growth.

The purpose of this mixed methods study was to gather and analyze data from teachers in City Public Schools who are currently being both observed and evaluated using the Stronge TEPES (2006). By using mixed methods, the qualitative and the quantitative data provided me with information in order to draw final conclusions. The survey included questions that asked participants to rate statements using the Likert scale in order to collect data regarding the implementation process, the potential for the process to create positive change, the quality of feedback, and future needs from the perspective of teachers in order to inform future practice and to improve administrator use of the Stronge TEPES (2006). Follow up interviews, as well as an analysis of summative evaluation scores from over a two-year period, were used to increase the value of the data, findings, and conclusions. The conclusions from this study may be used to inform

evaluation strategies that better meet professional growth which ultimately should improve student achievement.

The following research questions guided my study:

1. How do teachers in the City Public Schools perceive the implementation, specifically the ability to create positive change, of the Stronge Teacher Effectiveness Performance Evaluation System?

Data to answer this research question came from both the survey and the interviews. Participants had the opportunity to share their perspectives about the implementation of the Stronge TEPES (2006) through questions on the Likert scale. Additionally, those participants involved in the second collection of data through interviews were able to expand on their perceptions through semi-structured interview questions.

2. How do teachers in the City Public Schools perceive the quality of feedback they are receiving in observations and evaluations using the Stronge Teacher Effectiveness Performance Evaluation System?

The participants, all teachers, in this study shared their perceptions on the quality of feedback they have received on both observations and evaluations through the survey and interviews. They had the opportunity to discuss if the feedback they received had the potential to lead to professional growth and opportunities.

3. How do the summative evaluation scores of teachers in the City Public Schools demonstrate consistency in the use of the Stronge Teacher Effectiveness Performance Evaluation System?

Final summative evaluation scores, between zero and four, were analyzed for a random sampling of teachers. The scores from two consecutive years were compared to show an increase, decrease, or no change.

4. How can the Stronge Teacher Effectiveness Performance Evaluation System be used to shift the supervisory model in City Public Schools from evaluation to instructional supervision and professional development?

Through an examination and analysis of the three data sources, the current supervisory model was examined. The teachers shared their personal experiences with observation, evaluations, feedback, and implementation. These may be used to inform future practice.

Research Design: Mixed Methods

Mixed methods research is a mixture and combination of both qualitative and quantitative research (Creswell & Plano Clark, 2011). It is a way to collect, analyze, and mix both qualitative and quantitative research into one study (Creswell & Plano Clark, 2011). Johnson and Onwuegbuzie (2004) define mixed methods research “as the class of research where the researcher mixes or combines quantitative and qualitative research techniques, methods, approaches, concepts or language into a single study (pg. 17)”. As a newer form of research, it requires the researcher to have a clear understanding of both qualitative and quantitative research separately in order to properly and effectively combine the two to create one research study (Creswell & Plano Clark, 2011).

Using a mixed methods approach to this research provided an opportunity to have multiple perspectives when analyzing data because neither qualitative nor quantitative alone was enough (Creswell & Plano Clark, 2011). The topic of this study,

the Stronge TEPES (2006), provides both qualitative and quantitative feedback to teachers which sets the preference for a mixed methods study. It provided multiple viewpoints, both subjective and objective, which is appropriate when attempting to answer research questions (Creswell & Plano Clark, 2011). In this study, the perspectives came from multiple teachers in the City Public Schools, as well as various data forms (surveys, interviews, and summative evaluation score analysis). Additionally, because participants teach in various schools throughout the district, the study showed how the implementation process, the perceptions of its ability to create positive change, and the feedback provided to teachers varied from one school to another, as well as one evaluator to another. The survey did not include any identifying information about either the individual or school.

Research Design: Approach

Specifically, the convergent parallel design of mixed methods research was utilized in this research project. Creswell and Plano Clark (2011) discuss the primary purpose of the convergent parallel design as providing the researcher with the opportunity to obtain different, qualitative and quantitative data, but complimentary data on the same topic. Both the qualitative and quantitative data were collected simultaneously, a survey with Likert scale questions and follow-up interviews. By collecting both data sets during the same phase, I was able to identify patterns and themes in order to draw final conclusions. Schoonenboom and Johnson (2017) discuss the importance of timing in mixed methods; parallel design is when both components (quantitative and qualitative) are executed simultaneously. The final interpretations in this study came from a corroboration of findings within the study, in order to strengthen one another. Neither the

qualitative data nor the quantitative data alone would have provided a full understanding of the phenomenon, which required triangulation. Creswell (2014) discusses triangulation as a means to use more than one method to collect data on the same topic. This is a way of assuring the validity of research through the use of a variety of methods to collect data on the same topic, which involves different types of samples as well as methods of data collection. Additionally, the final collection of data, quantitative data, came from an analysis of summative evaluation scores. This analysis continued to corroborate the findings from the survey and the interviews.

The survey was given to teachers in the City Public Schools who have been evaluated using the Stronge TEPES (2006), and the interviews were completed with 13 willing teachers in the district. These conclusions helped to determine if there is a difference in perceptions among the teachers, which helped district administrators and observers become more aligned in practice.

Sampling Strategy

Criterion based sampling (Patton, 2002) was used to choose participants for the survey, as well as the participants in the summative evaluation score analysis. Patton (2002) describes criterion based sampling as a way to study all cases who meet predetermined criteria. In this study, the predetermined criteria were: participants must hold the position as a teacher in City Public Schools and each received at least one evaluation using the Stronge TEPES (2006). This sampling method was employed to ensure that all participants had the same prior experience before the study (Patton, 2002).

After completing the survey, the last question provided teachers with the opportunity to opt into the interview process of the study. Using simple random

sampling, a technique where a sample of participants is selected from a larger group (Fowler, 2009), participants were chosen to be part of the interview process. Each individual was chosen entirely by chance, and each participant had an equal chance of being included in the sample. Of the 23 teachers who volunteered to be interviewed, 13 were randomly chosen.

The summative teaching evaluation score analysis was completed by first utilizing criterion based sampling (Patton, 2002) and then simple random sampling (Fowler, 2009). Six teachers' summative scores were analyzed from each school for a total of 36 teachers across the district. The teachers' summative evaluation scores selected to be analyzed had no correlation to participation in either the survey or the interview. Only teachers with two consecutive evaluations scores were considered and then the participants were randomly selected.

Participants

There are currently 378 teachers in City Public Schools in grades PreK-12 all of whom have received professional development on how to use the Stronge TEPES (2006) either when it was initially implemented or when they began their employment. In order to be considered for this study, participants held a position as a teacher in City Public Schools and received at least one summative evaluation using the Stronge TEPES (2006). This sampling method was employed to ensure that all participants had the same prior experience before the study (Patton, 2002).

An email invitation was sent to each of the 378 teachers invited to participate in the study; 108 completed the survey through Qualtrics. Prior to beginning the survey, the participants self-selected into the study through an informed consent letter. After

completing the survey, teachers had the opportunity to opt into the interview process of the study. Twenty three teachers (of the 108 surveys completed) volunteered to be interviewed. The 13 interviewees were chosen using simple random sampling (Fowler, 2009). Additionally, six teachers' summative scores were analyzed from each of the district's six schools for a total of 36 teachers in the district.

The participants who completed the survey varied in both the number of years of service and the grade levels/subjects they teach. The only requirement was that they have been observed and evaluated using the Stronge TEPES (2006) a minimum of one time. The participants in the interviews also varied in those aspects because they volunteered to participate and then were randomly selected.

Setting

City Public Schools is an urban school community located in North Central, New Jersey. The district consists of four elementary schools, one middle school, and one high school. Each elementary school has students in grades Preschool-6. The middle school houses students in grades seven and eight, and the high school consists of students in grades nine to twelve. In addition, the district has an Alternative Center for Education, where approximately 125 students in grades 10-12 complete their coursework. Each grade level has approximately 270 students for a total population of approximately 3800 students. There are 328 teachers and 31 administrators in the district.

The demographics in City Public Schools have shifted drastically in the last decade to now include a diverse population of students that is 18% white, 40% African American, 35% Hispanic, 3% Asian, and 4% multi-racial. Approximately, 60% of the students in City Public Schools receive free or reduced lunch.

According to state test results two of the schools have been identified as in need of targeted assistance and support. Four of the six schools are classified as Title I Schoolwide programs, according to the NJ DOE.

Role of Researcher: Worldview

When completing a research study, the role of the researcher contributes to the worldview (Creswell, 2014), the beliefs and assumptions that inform a study. Creswell (2014) discusses the worldviews of research as meaning “a basic set of beliefs that guide action (pg. 6)”. Creswell & Plan Clark (2011) define it as “focuses on the consequences of research, on the primary importance of the question asked rather than the methods, and on the use of multiple methods of data collection to inform the problems under study (pg. 415)”.

Tashakkori and Teddlie (2003) suggest that pragmatism is the best worldview for mixed methods research, as it is problem centered and real world oriented. In this study, the research is focused on the perceived problem, that neither instruction nor student achievement have been positively impacted in City Public Schools since the adoption and the implementation of the Stronge TEPES (2006). This research will be real world oriented with the intention of shifting future practice for teachers and administrators.

Instrumentation and Data Collection

Using a convergent parallel design, quantitative and qualitative data were collected through a survey using a Likert scale and semi-structured interviews. In order to complete the data collection process, consent was received from the Superintendent of Schools in City Public Schools. Additionally, summative evaluation scores from six randomly selected teachers in each school were analyzed.

Survey. The survey instrument was available through an online tool, Qualtrics. The completion of the survey was completely voluntary, anonymous, and took approximately 15 minutes. This survey collected the qualitative data from a total of 108 teachers in the district. Though the survey was anonymous, with the hope of collecting honest data, there was a question directly related to the participant's position. However, participants were not asked to provide their name. In most cases because there was more than one teacher per grade level and subject, it would be difficult to identify the respondent.

The survey included statements that participants rated using a Likert scale. There were fifteen questions on the survey, as well as an opportunity for teachers to indicate their willingness to be part of the interviews. The first two survey questions asked about the teacher's professional career: number of years in district and current grade level. The next five questions asked about the participants' overall perceptions with regards to the Stronge TEPES (2006) regarding the following items: the quality of teacher observations, the impact of the implementation, the overall quality of the implementation, the ability to create positive change, and the improvement of student learning. Several questions asked about the participants' perceptions regarding the feedback received on observations. These questions asked if the feedback provided: identified strengths, identified areas in need of improvement, identified teaching practices, was valuable, led to learning about teaching, caused a positive impact on teaching, caused personal reflection, and created positive change. The final survey question asked participants about the impact annual evaluations had on professional development for teachers. (See Appendix A).

Interview questions. Using simple random sampling (Fowler, 2009), a technique where a sample of participants is selected from a larger group, participants were chosen to be part of the interview process. Each individual was chosen entirely by chance, and each participant had an equal chance of being included in the sample. After completing the survey, participants were asked if they were willing to be interviewed; only at that time were they able to provide their email address. Of all the 23 willing participants, 13 interviews were completed. This number of interviews provided sufficient qualitative data without resulting in data saturation.

Before beginning each interview, the purpose of the study was explained to the participants, as well as their individual roles. Each interview was conducted in person, either in a classroom or district office, for approximately 30 minutes. Using a responsive interview style, the standard open-ended questions were asked in order to ensure consistency in responses and data. Each of these standard interview questions was directly correlated to one of the research questions, as recommended by Rubin and Rubin (2012). Initially participants were asked to describe their current position in City Public Schools and how long they have been employed in the district. Then they were asked about their perceptions of the implementation of the Stronge TEPES (2006) and if its implementation has brought any change to the district. They were also asked about the preparedness level of both the individual teacher and the district as a whole. The interview questions also addressed their perceptions about the quality of feedback they receive on both observations and evaluations, as well as how this feedback has affected their instruction. The final interview questions asked about how either professional growth opportunities or supervision of teachers has changed since the implementation of

the Stronge TEPES (2006). Follow up questions were asked when clarification or further detail was needed. (See Appendix B). Interviews were audio-recorded with the consent of each participant and transcribed immediately following the interview (Rubin & Rubin, 2012)

Summative evaluation scores. At the end of each evaluation cycle (September to June), every teacher receives a summative evaluation score (AchieveNJ, 2013). These scores are a weighted calculation that combine teacher practice, student growth objectives, (SGOs), and student growth percentiles (SGPs) to give a teacher a score between zero and four (AchieveNJ, 2013). For non-tested teachers, a summative evaluation is weighted 85% of teacher practice and 15% of SGO results. For tested teachers, a summative evaluation is weighted 70% of teacher practice, 5% for SGPs, and 25% for SGOs (AchieveNJ, 2013). When using the Stronge TEPES (2006), teacher practice consists of six performance standards: professional knowledge, instructional planning, instructional delivery, assessment of/for learning, learning environment, and professionalism. Each of the performance standards is worth the same percentage of the total weight for teacher practice.

For this study, 36 teachers were selected. Initially, criterion based sampling (Patton, 2002) was used in order to ensure that each participant had two consecutive years of summative evaluation scores. After this, simple random sampling (Fowler, 2009) was used to choose six teachers from each of the six district schools. Each teacher's scores were analyzed to determine the difference (positive or negative) in growth over the two year period. By analyzing summative evaluation scores, the consistency in administrator practice using the Stronge TEPES (2006) was considered.

Analysis of Data

Mills (2003) discusses several ways to organize collected data in research such as coding the data, display findings, and data mapping. Coding the data, using labels to identify recurring themes, was the most appropriate way to organize data for this study. Coding is a way of indexing or categorizing the text in order to establish a framework of thematic ideas about it (Gibbs, 2007). When using Initial Coding and Pattern Coding, several preliminary findings and themes emerged. Using Initial Coding, as described by Saldana (2013), provided me an opportunity to determine what the data was showing rather than relying on my initial assumptions which would have created a limitation and bias in the study.

Using First Cycle Coding as a starting point, I began to see patterns emerge in the data. Saldana (2013) terms Initial Coding as breaking down qualitative data into discrete parts, examining these parts closely while comparing them for similarities and differences. With this in mind, Pattern Coding, which Saldana (2013) describes as creating a category label that identifies similarly coded data in order to organize them into themes, was used for the second cycle of coding. It gave me the ability to sort initial codes in order to identify emerging categories, explanations, or sets (Saldana, 2013). Using the second cycle of codes, I saw how patterns overlapped. These pattern codes provided me with a way to begin to understand preliminary findings and then draw overarching conclusions. In addition, I analyzed the results of my survey using percentages to determine the frequency of responses. By doing this, I was able to get a sense of what the majority of teachers have experienced. This provided me with ample information to draw final conclusions.

Mixed methods research was used because the purpose of my research and data collection was to identify the perceptions of teachers. Analysis of the qualitative and quantitative data occurred at the same time, as this was a convergent parallel study, which provided me with the opportunity to obtain different, qualitative and quantitative data, but complimentary data on the same topic (Creswell & Plano Clark, 2011). The conclusions for this research came from interpreting the data together, rather than separately or one at a time, because neither the qualitative data nor the quantitative data was more important.

Triangulation of Data

Denzin (1978) and Patton (1999) describe several ways to triangulate data, including methods triangulation, triangulation of sources, analyst triangulation, and theory triangulation. In this study, methods triangulation was used to provide credibility to the findings.

Creswell (2014) discusses triangulation as a means to use more than one method to collect data on the same topic. This is a way of assuring the credibility of research through the use of a variety of methods to collect data on the same topic, which involves different types of samples as well as methods of data collection. Methods triangulation is used to check the consistency of findings generated by different data collection methods. It is common to have qualitative and quantitative data in a study which elucidate complementary aspects of the same phenomenon. Where these data sources converge will be of great interest.

Conclusion

The goal of this chapter was to present a discussion of the methodology that guided this study with the intent of answering the research questions. The research

design, approach, sampling strategy, and worldview all contributed to the organization of the study. The description of the participants and setting provided an opportunity to further understand the need for the study's completion. Through instrumentation, data collection, and data analysis all led to the findings of this study, which will be detailed further in the following chapter.

Chapter 4

Findings

When President Obama created the competitive Race to the Top grant, there were four primary goals; one of these was for each state to create a system to recruit, develop, reward and retain quality educators (2009). The grant required states to reform teacher evaluation in the hopes of reaching this particular goal (2009). This was based on the belief that systematic evaluation practices would provide districts with information that would lead to better decisions about which teachers to continue to employ, in turn leading to greater student achievement.

After being awarded federal funding, the New Jersey Department of Education (NJDOE), under the direction of former Governor Chris Christie, agreed to this specific requirement in an attempt to reform public education. The 2012 amendment to TeachNJ, which included the AchieveNJ (2013) mandates detailed very specific changes to previous teacher evaluation systems (TeachNJ, 2012). One aspect of the TeachNJ (2012) regulations was that each district in the state had to choose an evaluation model which included a four-point rating scale for a teacher's final evaluation. This initiative was one measure utilized to begin tenure reform. The premise was that ineffective tenured teachers could be held accountable for both their professional practices and student achievement (TeachNJ, 2012).

Prior to this, teacher observations consisted mostly of a narrative description of a lesson accompanied by qualitative comments from a single observer, who may have been a supervisor, assistant principal, or principal. The same was true of the annual summative evaluations that teachers received. AchieveNJ (2013) included the addition of quantitative measures to something that was previously completely qualitatively. It

also mandated the inclusion of student achievement in each teacher's summative evaluation. A teacher's summative evaluation is based on a score, between zero and four, and combines teacher practice, student growth objectives, (SGOs), and student growth percentiles (SGPs) (AchieveNJ, 2013).

City Public Schools, a pseudonym given to protect the confidentiality of the district, in accordance with the mandates of AchieveNJ (2013) adopted the Stronge TEPES (2006) in 2013. Beginning that September, all teachers in the district received observations and summative evaluations using the systems, and they all received professional development on the requirements of the Stronge TEPES (2006). Prior to implementation, the administrators and observers in the district received training on how to complete both an observation and a summative evaluation to ensure inter-rater reliability.

The Stronge TEPES (2006), one of the NJ DOE approved systems, includes seven standards: professional knowledge, instructional planning, instructional delivery, assessment of/for learning, learning environment, professionalism, and student achievement. The four-point rating scale for this evaluation tool includes: ineffective, partially effective, effective, and highly effective. This mixed methods study examined teacher perceptions of whether the Stronge TEPES (2006) has been successfully implemented in the City Public Schools and if teachers felt they were receiving quality feedback that helped to guide and improve instruction.

This chapter describes the findings from data analysis in relation to the research questions that guided this study. It includes a brief description of the participant selection process, along with their synthesized experiences based on an in-depth analysis of a

structured survey, semi-structured interviews, and an analysis of individualized teachers' summative evaluation scores using the Stronge TEPEs (2006) over a two-year period. Surveys were anonymously administered through Qualtrics with the purpose of assessing the perception and feelings of the participants. Participants self-selected into the study through an informed consent letter. Completion of the survey took approximately 15 minutes, and 108 surveys were completed. The survey included 17 questions; 13 of the questions asked the participants to rate statements using a predetermined Likert scale, ranging from a 1, strongly disagree to a 5, strongly agree. The other four questions asked participants about their position, length in the district, and willingness to be part of the interview process.

Interviews were audio-recorded with the consent of each participant and transcribed immediately following the interview (Rubin & Rubin, 2012). Thirteen interviews were completed with the purpose of better understanding the participants' opinions and experiences. Participants were provided copies of the interviews for member checking and given an opportunity to correct or clarify responses in the transcripts. The transcripts were analyzed using two coding cycles that allowed for the identification of key ideas and repeating patterns that emerged as themes. When presenting data, each participant was provided a pseudonym in order to maintain confidentiality.

The purpose of this mixed methods study was to examine teachers' perceptions of the Stronge TEPEs (2006) with regard to its ability to affect positive change in instruction, student achievement, as well as provide professional growth opportunities for teachers based on individualized feedback in observations and on summative evaluations.

Both qualitative and quantitative data provided relevant information in order to draw final conclusions. By employing a survey that includes statements which are rated using the Likert scale, data were collected regarding the implementation process, the potential for the process to create positive change, the quality of feedback, and future needs from the perspective of teachers in order to inform future practice and to improve administrator use of the Stronge TEPEs (2006), as well as the potential for a shift of teacher supervision to one of professional development. Follow up interviews, as well as an analysis of summative evaluation scores over a two-year period, were used to increase the value of the data, findings, and conclusions.

Description of the Design

The convergent parallel design of mixed methods research was utilized in this research study (Creswell & Plano Clark, 2011). A convergent parallel design entails that the researcher concurrently conducts the quantitative and qualitative elements in the same phase of the research process, weighs the methods equally, analyzes the two components independently, and interprets the results together (Creswell & Plano Clark, 2011). The qualitative data and the quantitative data were independent of one another; however, both the qualitative and quantitative data were collected simultaneously. By collecting both data sets during the same phase, patterns were identified in order to draw final conclusions from emerging themes. The final interpretations in this study came from a corroboration of findings within the study, in order to strengthen one another. Neither the qualitative data nor the quantitative data alone provided a full understanding of the phenomenon, whether teachers perceive the Stronge TEPEs (2006) as an effective method of evaluation of teaching and learning. This will require triangulation.

Additionally, the final collection of data from the quantitative phase came from an analysis of summative evaluation scores of individual teacher practice in the district over a two-year period. These scores include a preponderance of evidence from observations, SGOs, SGPs, and documentation logs. These scores are averaged and provide a summative rating of the effectiveness of a teacher for the school year. This analysis was used to continue to corroborate the findings from the survey and the interviews in order to later lead to conclusions.

Creswell and Plano Clark (2011) discuss the primary purpose of the convergent parallel design as providing the researcher with the opportunity to obtain different, qualitative and quantitative data, but complimentary data on the same topic. The survey was given to teachers in the City Public Schools who have been evaluated using the Stronge TEPES (2006), and the interviews were completed with thirteen willing teachers in the district. I was able to triangulate the data collected from the teachers, analyze the teachers' summative evaluation score reports, and then provide final conclusions based on both.

Description of Participants

City Public Schools is an urban school community in New Jersey that has four elementary schools, one middle school, and one high school. There are currently 378 teachers in the district, spanning grades PreK-12, and each teacher has received professional development on how to use the Stronge TEPES (2006). Criterion based sampling, a way to study all cases who meet predetermined criteria (Patton, 2002), was used to invite teachers to participate in the study. In this study, the predetermined criteria were: participants must hold the position as a teacher in City Public Schools and each received at least one evaluation using the Stronge TEPES (2006). This sampling method

was employed to ensure that all participants had the same prior experience before the study (Patton, 2002).

Of the 378 teachers invited to participate in the study, 108 completed the survey, constituting a 28.5% return rate. After completing the survey, the last question provided teachers with the opportunity to opt into the interview process of the study. Using simple random sampling, a technique where a sample of participants is selected from a larger group (Fowler, 2009), participants were chosen to be part of the interview process. Each individual was chosen entirely by chance, and each participant had an equal chance of being included in the sample. Of the 23 teachers who volunteered to be interviewed, 13 were randomly chosen. The final question of the survey asked participants to provide both their name and email address if they were willing to participate in the follow-up interview.

The summative teaching evaluation score analysis was completed from randomly selected teachers in each of the district's schools. Six teachers' summative scores were analyzed from each school for a total of 36 teachers across the district. The teachers' summative evaluation scores selected to be analyzed had no correlation to participation in either the survey or the interview. They were randomly selected from the 378 teachers in the district. Simple random sampling, a technique where a sample of participants is selected from a larger group (Fowler, 2009), was utilized to choose the participants for this aspect of the study. What follows provides an analysis of the data findings from the online survey, the 13 face-to-face interviews, and the analysis of 36 summative evaluation scores.

Table 1 shows the number of years of experience of the teachers that responded to the survey. It demonstrates that teachers with a variety of experience participated in the survey portion of the study. However, the largest group of teachers responding to the survey, 51.72% (45), was those with 16 or more years of experience in City Public Schools. Only 19.54% (17) of the teachers had 5 or fewer years of experience in the district, which also indicates that 80.45% (70) are tenured teachers. This indicates that teachers have multiple years of experience using the Stronge TEPES (2006) and can speak to its effectiveness with some experience.

Table 1

Survey Participants' Total Years Teaching in City Public Schools

	Frequency	Percent
1 to 5 years	17	19.54%
6 to 15 years	25	28.73%
16 or more years	45	51.72%

Table 2 shows the current teaching assignments of the survey participants. This also demonstrates that a variety in experience will potentially provide multiple points of view of the participants in the study. The largest groups of respondents were teachers of students from Kindergarten to grade 4, elementary teachers. This group accounted for 44.83% (39) participants. The smallest group 4.6% (4) identified themselves as district staff, working in grades K-12.

Table 2

Survey Participants' Grade Level Teaching Assignment

	Frequency	Percent
Kindergarten to grade 4	39	44.83%
Grades 5 to 8	20	22.99%
Grades 9 to 12	24	27.59%
K-12	4	4.6%

The 13 interview participants included 4 K-4 teachers, 2 teachers in grades 5-8 and 7 teachers in grades 9-12. The 36 summative teaching evaluation scores were equally distributed among the 6 school buildings. There were 24 teachers in grades K-4, 6 teachers in grades 5-8, and 6 teachers in grades 9-12.

Quantitative Findings

Both the online survey results and the analysis of the summative evaluation scores provided quantitative data which were used to provide insight into understanding the perceptions of teachers using the Stronge TEPES (2006). The quantitative data is organized into four categories: (1) overall quality and impact, (2) the perception of implementation, (3) the quality of feedback, and (4) the change in summative evaluation scores. Survey questions provided data for the first three categories and the analysis of summative evaluation scores provided data for the final category of quantitative data. These categories can be correlated to not only the survey questions but also the research questions which guided this study.

Overall quality and impact. After answering initial questions about their experience as a teacher, participants were asked to rate the overall quality of the teacher evaluation process, including classroom observations, and to rate the overall impact of the Stronge TEPES (2006) on professional practices and instruction. For both questions, teachers used a rating scale of 1 (very poor quality or no impact) to 5 (very high quality or strong impact). The analysis that follows collapses the rating scale scores of 1 and 2 together, as well as 4 and 5 together, to provide a greater understanding of the data.

Table 3 shows that the majority of respondents felt the overall quality of the evaluation process, including classroom observations, was high. The overall quality means how successful the district overall has been when using the Stronge TEPES (2006). Classroom observations include visiting classrooms, observing instruction, and observing work in non-classroom settings (Stronge & Associates, 2015). A total of 51.72% (45) respondents rated the quality either a 4 or 5, which demonstrates a positive perception of the tool. It should also be noted the 32.18% (28) participants rated the overall quality of the evaluation process as neither poor nor high.

Table 3

Overall Quality of the Evaluation Process, including Observations

	Frequency	Percent
1 to 2 (poor to very poor quality)	14	16.09%
3 (neither poor nor high quality)	28	32.18%
4 to 5 (high to very high quality)	45	51.72%

While the majority of the participants felt that the overall quality of the evaluation process was high, Table 4 shows that the participants have a negative perception of the overall impact that the Stronge TEPEs (2006) has on their professional practices, pertaining to instruction and professionalism. This dichotomy potentially exists because either teachers have not changed their instructional practices or the administrators have not made changes to professional development opportunities that lead to instructional change instructionally. Of the responses, 39 teachers (44.83%) rated the statement either a 1 or 2 (little or no impact). This does not indicate that the impact negatively affected professional practices, but rather did not impact the participant professionally. However, it should also be noted that 24 teachers (27.59%) felt that the Stronge TEPEs (2006) had a strong (positive or negative) impact on their professional practices.

Table 4

Overall Impact of the Stronge TEPEs on Professional Practices

	Frequency	Percent
1 to 2 (little or no impact)	39	44.83%
3	24	27.59%
4 to 5 (strong impact)	24	27.59%

These findings are significant because the goal when implementing the Stronge TEPEs (2006) in the City Public Schools was to not only comply with the AchieveNJ

(2013) mandates but to also begin to provide higher quality instruction to students in order to see a positive impact on student achievement.

Perception of implementation. After the questions about the overall quality and overall impact of the Stronge TEPES (2006), the survey included several items that correlated to the participants' perception of the implementation and its ability to affect change in the district. The district anticipated the change would lead to improvement of student learning, the reflection on teaching practices, and the implementation of high quality instruction. For each of these questions, teachers used a rating scale of 1 (very poor quality) to 5 (very high quality).

Teachers were initially asked to rate the overall quality of the implementation of the Stronge TEPES (2006) in City Public Schools. The question did not explicitly dictate to the participants what was meant by implementation process. However, these results indicate that teachers responded based on the initial professional development received by staff, as well as the first completed evaluation cycle using the Stronge TEPES (2006) system in the district. Table 5 shows that a large number of participants, 30 (34.48%), had a neutral view of the implementation process. In addition, 40 participants (45.98%) felt that the overall quality of the implementation of the Stronge TEPES (2006) was high. The implementation of the Stronge TEPES (2006) came in the form of professional trainings for all staff. Observers were provided several days of professional development on how to observe a class, including ensuring inter-rater reliability. Teachers received two days of professional development, provided by members of the Stronge TEPES team, followed by meetings led by building administrators. These occurred throughout the first year the system was implemented.

Table 5

Overall Quality of the Implementation of the Stronge TEPES

	Frequency	Percent
1 to 2 (poor to very poor quality)	17	19.54%
3	30	34.48%
4 to 5 (high to very high quality)	40	45.98%

The next series of questions asked teachers to use a five point Likert Scale to rate a series of statements. The same Likert Scale (1= Strongly Disagree and 5= Strongly Agree) was used throughout the remainder of the survey. The participants were first asked to rate whether the implementation of the Stronge TEPES (2006) had the ability to create positive change in the district. The change being asked about is the improvement of student learning, the reflection on teaching practices, and the implementation of high quality instruction. The largest percentage of teachers neither agreed nor disagreed with the statement. Table 6 shows the responses with 42.53% (37) teachers choosing 3 as their answer. It can also be noted that 33.33% (29) responses either strongly disagreed or disagreed. The positive change being alluded to in this statement is broken down in the next several questions. For example, the questions ask about the improvement in student learning and the reflection of teachers on their own professional practices. Even further, this can be related back to the goals of AchieveNJ (2013) and TeachNJ (2012) to improve instruction and increase student achievement throughout the state. Both of these are areas

that are included in determining whether the summative evaluation results have the ability to create a positive change in the district.

Table 6

The Ability to Create Positive Change in the District

	Frequency	Percent
1 to 2 (disagree to strongly disagree)	29	33.33%
3 (neither disagree nor agree)	37	42.53%
4 to 5 (agree to strongly agree)	21	24.14%

The final survey question regarding the Stronge TEPES (2006) implementation and its ability to create change asked teachers to rate a statement, “I think teacher evaluation improves the learning of my students”, using the 5 point Likert scale. The change being referred to in the statement was an increase in student achievement and learning as a direct outcome of the implementation. Table 7 demonstrates that 41.39% (36) of responses strongly disagreed or disagreed with the statement. In addition, 29.89% (26) participants were neutral about the Stronge TEPES (2006) being able to create positive change in the district because they neither disagreed nor agreed with the statement in this survey question.

This question demonstrates that the improvement of learning for students was not the positive change that participants had initially anticipated. It is possible that

teachers in the district have not taken the feedback provided in classroom observations and the feedback and the results of summative evaluations into account when planning for future instruction. It is possible that through improved instruction the implementation and use of the Stronge TEPES (2006) could lead to increased student achievement and learning.

Table 7

Teacher Evaluation Improves the Learning of my Students

	Frequency	Percent
1 to 2 (strongly disagree to disagree)	36	41.38%
3 (neither disagree nor agree)	26	29.89%
4 (agree to strongly agree)	25	28.73%

The responses and ratings of this statement demonstrate that the participants see a lack of connection between teacher evaluation using the Stronge TEPES (2006) and student learning.

Quality of feedback. Survey results from several questions provided data to determine if the teachers perceived the quality of feedback they are receiving in observations and evaluations using the Stronge TEPES (2006) as valuable and useful. Each of the survey questions related to the quality of feedback teachers perceived they

were receiving from administrators on both observations and summative evaluations.

Teachers used the same Likert scale, as referenced previously, to rate a series of statements.

The first statement teachers were asked to rate was “The written feedback that I was provided after formal observations helped me to identify strengths in my teaching.” The majority of teachers, 52.87% (46) responses, agreed with this statement indicating that observations and evaluations provided evidence regarding teachers’ strengths in the classroom. Twenty one teachers, 24.14%, either disagreed or strongly disagreed with the statement. This indicates that teachers had the perception that they were not being provided with evidence that identified strengths in their instruction and pedagogy. Table 8 displays the distribution of the responses.

Table 8

Written Feedback Helped Identify Strengths in Teaching

	Frequency	Percent
1 or 2 (strongly disagree or disagree)	21	24.14%
3 (neither disagree nor agree)	20	22.99%
4 or 5 (agree or strongly agree)	46	52.87%

The next statement asked teachers to rate the statement, “The written feedback that I was provided after formal observations helped me to identify areas needing improvement in my teaching” using a five point Likert scale. As with the previous statement regarding the teachers being able to identify strengths in their instructional practices through the feedback received on classroom observations, the responses for this question demonstrated 52.87% (46) agreed with this statement and were able to identify areas needing improvement based on the feedback received on classroom observations. This indicates that teachers had the perception that they were being provided with a preponderance of evidence that identified areas of weakness in their instructional practices and pedagogy. Table 9 displays the distribution of responses for this question. It shows an increase in the number of responses at each level moving from strongly disagree to agree. There was a decrease in the number of responses from agree to strongly agree.

Table 9

Written Feedback Helped Identify Areas Needing Improvement in Teaching

	Frequency	Percent
1 or 2 (strongly disagree or disagree)	19	21.84%
3 (neither disagree nor agree)	22	25.29%
4 or 5 (agree or strongly agree)	46	52.87%

Both of the previous statements demonstrate that teachers have confidence in the observers' ability to provide evidence regarding the strengths and weaknesses in their pedagogy. Being able to identify both strengths and weaknesses in instruction will provide teachers with the necessary information to make future decisions regarding instructional practices. This also demonstrates that observers do not only speak to either the strengths or the weaknesses but rather both, meaning that they are not slanting observations in one manner or the other.

The next survey question asked teachers to rate the statement, "The written feedback that I received after formal observations was valuable." As noted in Table 10, the largest number of responses, 60.02% (53) agreed with this statement. It can be summarized that the evidence being provided to teachers to help identify both strengths and weaknesses in pedagogy was being viewed and interpreted as valuable, meaning it helped teachers to identify necessary pedagogical changes. There was also 21.84% (19) teachers who rated this statement as strongly disagree or disagree, further validating its value.

Table 10

Written Feedback after Formal Observations was Valuable

	Frequency	Percent
1 or 2 (strongly disagree or disagree)	19	21.84%
3 (neither disagree nor agree)	15	17.24%
4 or 5 (agree or strongly agree)	53	60.02%

When participants were asked to rate the statement, “The written feedback that I received after formal observations was meaningful in terms of identifying teaching practices I used”, 58.62% (51) either strongly agreed or agreed with it. Table 11 shows that the smallest percentage of responses 13.80% (12) were in the strongly disagree or disagree category. This statement asked participants to rate if the feedback assisted teachers in identifying the pedagogical practices used during the lesson(s) observed. This statement did not ask about the quality of the pedagogical practices observed in a lesson. These four survey questions asked participants to classify the feedback they received in different categories rather than ask the participants about the actions they have each personally taken after receiving feedback on multiple observation reports throughout the year

Table 11

Written Feedback after Formal Observations was Meaningful, Identifying Teaching Practices

	Frequency	Percent
1 or 2 (strongly disagree or disagree)	12	13.80%
3 (neither disagree nor agree)	24	27.59%
4 or 5 (agree or strongly agree)	51	58.62%

These four statements asked the participants to rate the feedback they were receiving. They were not asked to rate what they changed in their practices because of the feedback. In all four statements, the majority of participants either agreed or strongly agreed with the statements. This demonstrates that the teachers felt a high level of confidence in the observers' ability to provide quality feedback. Administrators and observers are using the rubric to identify teaching practices, strengths, and weaknesses for teachers.

Effects of feedback on instruction. The next several questions asked the participants to consider the effects that the feedback they received had on their practice and professional decisions, which differs from merely identifying practices, strengths, or areas for improvement as the previous questions did. The data indicates that the feedback is provided; however, these questions demonstrate whether the feedback is causing any changes or shifts in teacher practice.

Initially teachers were asked to rate the statement, "The written feedback that I received after formal observations led to learning about teaching on my part." This statement infers that teachers took personal responsibility to learn about best practice and pedagogy based on the feedback from observers. Table 12 shows that there was nearly an equal distribution of responses; each category (from 1-5) received between 11 and 25 responses. However, the largest number of responses 41.38% (36) agreed or strongly agreed that they utilized the feedback after receiving it and took personal responsibility and initiative to learn about teaching based on the evidence and feedback provided. Learning about teaching should lead to an improvement of instructional strategies and potentially an increase in student achievement, both goals of AchieveNJ (2013) and teacher evaluation reform.

Table 12

Written Feedback after Formal Observations led to Learning about Teaching

	Frequency	Percent
1 or 2 (strongly disagree or disagree)	28	32.18%
3 (neither disagree nor agree)	23	26.44%
4 or 5 (agree or strongly agree)	36	41.38%

Table 13 is a depiction of the responses about the statement, “The written feedback that I received after formal observations has had a positive influence on my teaching practices.” Different from the previous statement, which asked about a respondent’s personal learning about teaching, this statement seeks to determine if the feedback led to a positive influence on professional practices in the classroom. This takes the previous statement a little further in the area of personal accountability and responsibility. Rather than just learning about teaching, participants were asked if they put this learning in to practice when providing instruction. It asks if the participants were willing to make changes based on the feedback they have received on classroom observations. It infers a trust that the feedback is reliable based on the knowledge of the person providing it. Even more participants, 48.27% (42), indicated that the feedback provided to them through formal observations had a positive influence on their pedagogy. This is important because it demonstrates the participants’ belief that they are making

positive changes to their practices because the Stronge TEPEs (2006) has been implemented in the district.

Table 13

Written Feedback after Formal Observations had Positive Influence on Teaching Practices

	Frequency	Percent
1 or 2 (strongly disagree or disagree)	21	24.14%
3 (neither disagree nor agree)	24	27.59%
4 or 5 (agree or strongly agree)	42	48.27%

Teachers were also asked to rate the statement, “The written feedback that I received after formal observations caused me to reflect on my teaching.” This statement asked if teachers were internalizing the feedback in a way that caused them to think about their pedagogical choices. This statement did not ask if they made changes, but rather reflected and contemplated their actions. Of the responses, 54.02% (47) either agreed or strongly agreed with the statement, as seen in Table 14.

Table 14

Written Feedback after Formal Observations caused Reflection on Teaching

	Frequency	Percent
1 or 2 (strongly disagree or disagree)	20	22.99%
3 (neither disagree nor agree)	20	22.99%
4 or 5 (agree or strongly agree)	47	54.02%

The final survey question in this series asked teachers to rate the statement, “The written feedback that I received after formal observations has led or will lead to positive changes in my teaching practices.” This question was a culmination of this series of questions which led to participants contemplating the effects on learning about pedagogy to influence on instruction to reflection of practice to ultimately making pedagogical changes. Different from asking teachers about the overall impact of the Stronge TEPES (2006) on the district, this question asked participants to contemplate if the feedback provided to them through observations from multiple observers led to positive changes in teaching for each individual. Table 15 shows that 31.03% (27) neither agreed nor disagreed with the statement. However, it should be noted that only 21.84% (19) either disagreed or strongly disagreed. With 47.12% (41) either agreeing or strongly agreeing, it can be concluded that overall the participants thought the feedback not only could but did lead to positive changes in teaching practices. The percentage of teachers who responded that they reflected on their teaching was higher than those that responded that

they made positive changes in their teaching. While personal reflection can be difficult, making actual changes to practice requires more time.

Table 15

Written Feedback after Formal Observations has led to Positive Changes in Teaching Practices

	Frequency	Percent
1 or 2 (strongly disagree or disagree)	19	21.83%
3 (neither disagree nor agree)	27	31.03%
4 or 5 (agree or strongly agree)	41	47.12%

The final survey question using the Likert Scale asked teachers to rate the statement, “My annual evaluation leads to getting professional development and support for my improvement.” This statement shows the teachers’ perception regarding follow up after observations and evaluations from the district administrators in the areas of professional development for teachers and various supports for improvement for teachers. In contrast from the previous several questions, this statement did not ask about the teachers’ actions or thoughts based on feedback, but instead asked the participants to rate a statement about what the observers in the district offered following the observations and summative evaluations. Table 16 shows the distribution of data. The largest percentage of participants, 42.52% (37), either disagreed or strongly disagreed with this statement. This shows a lack of follow through from the administrators in the district.

The observations and summative evaluations have not bridged a gap to professional development and teacher growth.

Table 16

Annual Evaluation leads to Professional Development and Support for Improvement

	Frequency	Percent
1 or 2 (strongly disagree or disagree)	37	42.52%
3 (neither disagree nor agree)	25	28.74%
4 or 5 (agree or strongly agree)	25	28.74%

It is evident through the analysis of the survey results that teachers have the perception that they are receiving quality feedback, about both strengths and weaknesses in their instruction, and that this feedback has caused additional personal learning, reflection, and individual changes in teaching practices. However, the feedback has not led to changes in teacher supervision and support from administrators in the area professional development. It should also be noted that on each of the survey questions about feedback between 17.24% (15) and 31.03% (27) neither disagree nor agree with the statements indicating these participants have not been impacted either positively or negatively by the feedback they are receiving on observations.

The change in summative evaluation scores. The next set of quantitative data collected was obtained from summative teaching evaluations scores over a two-year

period in order to determine if there was any consistency in the scoring of summative evaluations. In each of the six schools in the district, six teachers were randomly selected. These teachers did not have to participate in either the survey or the interview. However, the criterion was that the teachers' whose scores were chosen had to be district employees for a minimum of two consecutive school years, 2016-17 and 2017-18 in order to ensure that a comparison could occur.

Table 17 lists the 36 teachers, anonymously and without any description, along with their two consecutive summative evaluation scores and the difference between the two. The table is organized by the difference in score. The first set of teachers saw no change followed by the teachers who saw a decrease in summative teaching evaluation score. These teachers are listed in descending order. The last set of scores is the teachers who saw an increase in the summative teaching evaluation scores and are also listed in descending order.

Of the 36 teachers included in this data set, 36% (12) received the same score from the first year to the second year. In addition, the table includes a listing of the six schools, by number rather than name, to provide further analysis and information. All the teachers included in this data set from School 1 saw no difference in their summative evaluation score over the two-year period.

In addition, the two greatest decreases in scores over the two-year period showed a difference of .42 for Teacher 13 and .29 and Teacher 14. These teachers were from two different schools in the district. Not only were these the two greatest decreases in scores, but they were also the greatest change (either increase or decrease) in scores over a two-year period, either negative or positive. The two greatest increases in scores were .21 for

Teacher 23 and .19 for Teacher 24. Similarly, these teachers were from two different schools in the district. It should be noted that 27.78% (10) teachers' scores decreased over the two-year period, and 37.83% (14) teachers saw an increase in their final evaluation score.

Table 17

Summative Evaluation Scores

	School	2016-17	2017-18	Difference
Teacher 1	School 1	3.72	3.72	0
Teacher 2	School 1	3.22	3.22	0
Teacher 3	School 1	3.25	3.25	0
Teacher 4	School 1	3.58	3.58	0
Teacher 5	School 1	3.86	3.86	0
Teacher 6	School 1	3.58	3.58	0
Teacher 7	School 2	3.67	3.67	0
Teacher 8	School 3	3.43	3.43	0
Teacher 9	School 4	3.86	3.86	0
Teacher 10	School 4	3.15	3.15	0
Teacher 11	School 4	3	3	0
Teacher 12	School 5	3.28	3.28	0
Teacher 13	School 3	3.33	2.91	-.42
Teacher 14	School 2	3.43	3.14	-.29

Table 17 Continued

Summative Evaluation Scores

	School	2016-17	2017-18	Difference
Teacher 15	School 3	3.29	3.09	-.20
Teacher 16	School 2	3.43	3.25	-.18
Teacher 17	School 2	2.86	2.72	-.14
Teacher 18	School 3	3.18	3.11	-.07
Teacher 19	School 2	3.24	3.18	-.06
Teacher 20	School 5	3.67	3.62	-.05
Teacher 21	School 6	3.8	3.76	-.04
Teacher 22	School 6	3.43	3.4	-.03
Teacher 23	School 6	3.09	3.3	+.21
Teacher 24	School 4	2.97	3.16	+.19
Teacher 25	School 5	3.42	3.58	+.16
Teacher 26	School 4	3.75	3.9	+.15
Teacher 27	School 5	3.43	3.58	+.15
Teacher 28	School 6	3.43	3.58	+.15
Teacher 29	School 6	3.18	3.32	+.14
Teacher 30	School 3	3.55	3.69	+.14
Teacher 31	School 4	3.15	3.29	+.14
Teacher 32	School 3	3.16	3.28	+.12

Table 17 Continued

Summative Evaluation Scores

	School	2016-17	2017-18	Difference
Teacher 33	School 6	3.09	3.21	+.11
Teacher 34	School 5	3.29	3.36	+.07
Teacher 35	School 2	3.08	3.15	+.07
Teacher 36	School 5	3.33	3.36	+.03

The minimal changes that occurred in the scores over a two-year period indicate two things: the administrators in the district are potentially using the previous year's summative teaching evaluation score as a guide in the current year's summative evaluation and/or there are no significant changes in instructional practice of teachers to have a real impact. This data reiterates and confirms the survey data that teachers are receiving feedback but are not necessarily making the necessary shifts in professional practices.

The four categories found in the quantitative data (overall quality and impact, the perception of implementation, the quality of feedback, and the change in summative evaluation scores) led to a better understanding of the teachers' perceptions about the implementation, adoption, and usage of the Stronge TEPES (2006). In combination with the qualitative findings, these four categories led to conclusions and answers to the primary research questions that guided this study.

Qualitative Findings

Using simple random sampling, a technique where a sample of participants is selected from a larger group (Fowler, 2009), participants were chosen to be part of the interview process. Once completing the anonymous survey, teachers were able to indicate on the last question if they were willing to participate in an interview. It was not until this time that they were prompted to include their email addresses. Twenty three teachers volunteered to be part of the qualitative data collection process; 13 of those 23 teachers were randomly selected using simple random sampling (Fowler, 2009). Each individual was chosen entirely by chance, and each participant had an equal chance of being included in the sample.

Before beginning each interview, the purpose of the study was explained to the participants, as well as their individual roles. In addition, participants signed an informed consent form. Each interview was conducted in person, either in the individual teacher's classroom or a district office, and lasted approximately 30 minutes. Using a responsive interview style (Rubin & Rubin, 2012), the standard open-ended questions were asked in order to ensure consistency in responses and data. Each of these standard interview questions was directly correlated to one of the research questions, as recommended by Rubin and Rubin (2012). Additionally, the interviewees were asked probing questions to clarify and expand upon initial responses given to the standard interview questions.

Mills (2003) discussed several ways to organize collected qualitative data in research such as coding the data, display findings, and data mapping. Coding the data, using labels to identify recurring themes, was the most appropriate way to organize data for this study. Initial coding and pattern coding provided me an opportunity to determine

what the data was showing rather than relying on my initial assumptions which would create a limitation and bias in the study.

Using First Cycle Coding as a starting point, I was able to begin to see patterns emerge in the data. Initial Coding is described by Saldana (2013) as breaking down qualitative data into discrete parts, closely examining them, and comparing them for similarities and differences, and was utilized because it provided an opportunity to determine what the data were showing rather than what I wanted it to show me based on assumptions. With this in mind, pattern coding, described as creating a category label that identifies similarly coded data and organizes it into themes, was used for the Second Cycle Coding. It gave me the ability to sort initial codes to identify emerging categories, explanations, or sets (Saldana, 2013). These pattern codes provided me a way to begin to understand preliminary findings and then draw overarching conclusions.

Thirteen interviews were completed to collect qualitative data. Each participant was given a pseudonym in order to maintain anonymity. Table 18 provides the list of the teachers interviewed, their position and the number of years they have taught in the district. This information was collected through the initial interview question.

Table 18

Interview Participants

Teacher Name	Position	Years Teaching in District
Adam	Elementary Special Education	11
Barbara	High School English	5
Charles	High School Special Education	5
David	Middle School Math	2
Elizabeth	Kindergarten	23
Francesca	High School English	12
Gabrielle	High School English	23
Heidi	Theatre	5
Isabella	Middle School Language Arts	11
Jennifer	Elementary	4
Kim	Elementary	18
Larry	High School World Language	9
Mary	High School Special Education	30

After completing both the first and second cycle of codes, there were five major themes that emerged through the interviews: (1) conflicting perceptions of the implementation of the Stronge TEPES, (2) the potential for change, (3) administrator use of cookie cutter feedback, (4) instruction unaffected by the Stronge TEPES, and (5) lack

of guidance provided for professional development. The interview questions were geared toward understanding and gathering data about the teachers' perceptions, as well as gathering data to answer each research question.

Conflicting Perceptions of the Implementation of the Stronge TEPES

The first theme that emerged was the conflicting perceptions of the implementation of the Stronge TEPES (2006) in City Public Schools. Some of the participants provided data demonstrating that they perceived the implementation as successful, while others spoke to the implementation as being a challenge. Based on the responses, the perception each participant has of the implementation of the Stronge TEPES (2006) comes from the professional development provided by the district, as well as the initial use of the system by administrators and teachers. When asked about the implementation, some participants' responses demonstrated a confidence in the district's professional development and training of both the teaching staff and the district administrators, who act as the observers.

Successful implementation. Of the thirteen interviews, several responses described the implementation being successful and leading to an understanding of using the Stronge TEPES (2006), which is the first sub-theme that emerged. Mary commented, "Implementation was very good. It was well-explained." Isabella went a bit further to state,

I remember that there was training, and that training was good; very good in fact, especially in terms of how the documentation log was going to be completed. And I think the different parts of the Stronge model were explained pretty well. So I

feel like the roll out for adopting it was pretty well supported in terms of PD and things like that.

It is evident that Isabella felt that the teachers in her professional development workshop were provided with adequate information on how to use the system.

Some teachers discussed how they were given an opportunity to learn prior to implementation or even after the initial professional development. Adam said,

As far as the implementation of it, in the literature and my reading about it, it seems that it was implemented correctly. I mean, I have seen observations and evaluations from people in the districts that do the Danielson Model and it's all the same stuff.

Adam's response also indicates that he continued to educate himself about the Stronge TEPES (2006) after the initial implementation to ensure that he had sufficient knowledge of the system and how to implement it. Gabrielle put herself in a position to learn about the Stronge TEPES (2006) prior to the implementation throughout the district. She stated,

We had some full faculty meetings, where the principal explained to us the decision making process. I was fortunate in that I volunteered to be on the DEAC that looked at various programs and the requirements of each program. We did have a group of teachers who were willing and interested to look at the whole process. But for the rest of the faculty if you weren't one of the volunteers it was just sort of explained to you that this was determined to be the best program and then we began implementing it.

Gabrielle's response demonstrates that without being part of the initial committee to choose the program, she believes the initial implementation training was minimal. If additional learning led to proper usage and understanding, the initial professional development may not have been sufficient for all.

The perception of the implementation did not seem to change based on when the teacher was hired. Charles was not in the district during the 2013 school year, when the Stronge TEPEs (2006) was initially implemented. However, he stated,

Well I thought you guys were very thorough in the way you explained it. My first year here we were still just doing Stronge, I believe. And then I think my second year, you guys gave us like a little refresher on Stronge Plus, I believe it was.

So a big part of the New Teacher Orientation was explaining how we will be observed and evaluated. The second year we went to Stronge Plus, which pretty much was the same thing as Stronge, but with a couple tiny tweaks to it. There wasn't really anything that affected us in like a major way, but it still explained in detail.

Kim, who has been a teacher in the district for 18 years, was present for the initial implementation with the entire faculty, which differed from Charles who received professional development during New Teacher Orientation. Despite this contrast, Kim also felt the implementation was a positive learning experience. She described,

It was a while back. Well, we were told that this was a statewide initiative, and it was supposed to, I think, level the playing field for teachers so that we're all rated against the same system. It was, I think, to make it fairer because this evaluation system leaves some, but very little room, for the administrator to be subjective.

It's more of an objective evaluation system. Years ago, when there was nothing to really follow, it was basically whoever observed you and if they liked you, you got great but if you and this person didn't get along, maybe not so great. But, there was nothing official to monitor this. As far as the teacher was concerned, I never saw like a rubric or anything like that. With this evaluation system, I think it's more upfront for the teacher. You can see exactly what criteria are being evaluated. We learned all of this right away.

Kim described the implementation of the Stronge TEPES (2006) as a way to create equity in the way in which teachers are evaluated. It took away the subjective nature of both observations and evaluations.

These participants felt that the successful implementation of the Stronge TEPES (2006) was due to the initial professional development and training offered to the district staff. They had a level of confidence in their personal ability to use the system, as well as confidence in the administrators completing the observations and evaluations.

A challenging implementation. A few of the other participants who were interviewed had a contrasting opinion of the implementation and the professional development they received prior to using the system. These teachers had a very negative perception that they shared and felt that the implementation had many challenges thus leading to a challenging implementation, the second sub-theme that emerged. For example, Francesca discussed the changes that occurred even after the implementation. Based on her response, the perception was that the district was not fully prepared to implement and had to make changes based on use and comfort level. She said,

I would say that initially it was somewhat cumbersome, the way it was rolled out. A lot of paperwork and then there were some changes along the way that presented a bit of a learning curve before I think we were comfortable using it. I guess that would be all I would say about implementation.

This response indicates that the initial implementation did not necessarily prepare teachers to use the system.

Larry perceived the implementation and the role that administrators played in the implementation as very negative. He spoke to the lack of direction and lack of ability to answer questions or concerns.

It was not explained well when it was brought to us. And I'm only speaking for my school, the high school. I don't feel that there was a roadmap for us to follow, number one. Everything was sort of just thrown at us, and we were expected to discern what each step involved. That created a lot of skepticism and quite a bit of animosity. Again, this is my complaint, that what happens, and it's the first rule of acting, know your audience. It never seems as if the audience is taken into consideration. In this case, the audience was that teachers. You have to know who you're talking to, and how you communicate with them. By knowing them as well as you do as an administrator, you should know what their concerns would be, and you should target those concerns. That did not happen so Stronge got off on a very bad foot, I think.

It is evident through this response that Larry did not have confidence in the administrators' ability to roll out the system and explain it well to the teachers.

Elizabeth spoke to the implementation not only in terms of the Stronge TEPES (2006) in City Public Schools but also the way in which the NJ DOE and the government presented the AchieveNJ mandates (2013). She said,

Well, my initial perception was, okay, this is something else we have to do because the state department is shoving it down our throats. I think the district did their best to try and train us initially, but they couldn't change what the state demanded. I think they tried to do a by-the-book kind of implementation.

Her response acknowledged that what the district provided may have been adequate it was the fact that it was being forced which caused a concern. The mandates created a lack of buy-in and support.

As with any major initiative, there was varying understanding and comfort with the change. However, the perception of the implementation was not consistent among the participants. Each participant spoke directly to how the Stronge TEPES (2006) was rolled out to the teachers, but their recollection of the professional development and how they were able to move forward using the Stronge TEPES (2006) differed. The lack of consistency in the implementation has not led to consistent usage or understanding. Similar to the quantitative data, more participants felt that the implementation of the Stronge TEPES (2006) was successful in the district.

The Potential for Change

The second theme that emerged from the interviews was that the implementation of the Stronge TEPES (2006) had the potential to bring about change in the district. Two subthemes also emerged. Participants spoke directly to two specific changes: an increase in paperwork with a decrease in quality instruction and an increase in accountability of

instruction; both of these changes occurred because of compliance rather than a desire for improvement. The implementation of a four-point evaluation system was a requirement under AchieveNJ (2013). The mandates dictated timelines and narrowed the choices that the districts had. Through the interviews, it was evident that participants felt that the district, both administrators and teachers, were making changes because they felt they had to and not necessarily because they intrinsically believed in the need for a change.

Increase in paperwork with a decrease in quality instruction. The change in the evaluation system also brought with it an increase in paperwork with a decrease in quality instruction for both the teachers and the administrators. When speaking to this change, the participants most often demonstrated a negative perception of its impact on the district.

Adam spoke directly to the increase of paperwork and time that the Stronge TEPES (2006) has had on the teachers in the district and took it a step further when he spoke of the time implementation took from quality instruction.

I think the district when they first implemented it, I think it was fear. A lot of teachers were afraid of it like oh my god it's going to be this and we have to do this. But now we know, if you do your job, it doesn't affect you. Because it just comes down to doing your job. The only thing is there was, it seems, like there was a lot more paperwork and you had to go to more meetings and it seemed like you had to jump through more hoops. The paperwork takes time away from what we should be doing. This isn't really about teaching because we aren't really teaching with this. We are doing paperwork and a lot of it.

While the purpose of the Stronge TEPES (2006) is to improve instruction through evidence, Adam concluded that the system and its increase of paperwork do not provide teachers with the opportunity to improve practice due to a lack of time.

Elizabeth spoke negatively about the impact the Stronge TEPES (2006) has had on the teachers and administrators due to the amount of paperwork. As a teacher, she sees the administrators having less time to be instructional leaders because of the additional work needed to be compliant with both the guidelines for Stronge TEPES (2006) and AchieveNJ (2013). She said,

I think it's created far too much paperwork on the administration end and far too much paperwork on our end. That documentation log is ridiculous. It's like busy work for teachers. We're not supposed to give busy work to our kids, but we're given that amount of busy work. And we all, I think I can speak for most people, feel that if you want to see what going on in our classroom, come in our classrooms. Why are we wasting paper and ink to print pictures and write up things that we're doing? In the same breath, we're being told, "Don't use the copier. Don't do this. Don't do that. We're running out of money, blah, blah, blah". But then, the printer is off the hook when everyone is getting ready to finish their documentation log. It's silly.... As far as the evaluation goes, I think there's a lot of paperwork. Like, when you have your announced observation, you have to do this whole pre-conference thing, and then you have to do the post conference thing. It's like, come on. We're all professionals. I think it could be culled down into something simple, maybe what the objective of the lesson is that you're going to observe. But, I mean, that big, long form in *My Learning Plan*? You're

repeating things over and over again. It's silly. And again, it's another thing you have to do...

Francesca reiterated the same sentiment and perception when she spoke about the only change being in the amount of paperwork created rather than anything substantial changed in her instruction. This sentiment is not aligned to the overall goal of either AchieveNJ (2013) or the Stronge TEPES (2006). Both are meant to provide teachers with evidence of their instructional strengths and weakness. She stated,

I don't think it created a change at all other than the amount of, again, paperwork that we have to do to create it. I think that it doesn't necessarily, for me, inform my teaching. I think I had an understanding of my teaching outside of preparing this documentation log, and I don't know that it necessarily gave others more insight to my actual teaching as opposed to my ability to sort of put together a presentation of my teaching.

Each of these teachers spoke about the completion of the Stronge TEPES (2006), no matter the element, as being done in order to be compliant.

This increase in paperwork has not resulted in an increase in student achievement, as both AchieveNJ (2013) and the Stronge TEPES (2006) hope for. With teachers and administrators working toward compliance, there is no support for this initiative. This lack of support has led to mediocrity in the area of teacher evaluation. This qualitative data reiterates the quantitative data, which demonstrated that participants saw little to no impact on professional practices or student learning.

Increase in accountability for instruction. Another change that emerged when discussing the changes that occurred from the implementation of the Stronge TEPES

(2006) was the increase in instructional accountability teachers were experiencing. They spoke to higher instructional expectations and having to be aware of what's occurring in the classroom in a more diligent way. However, the perception was that this was occurring for the sake of compliance rather than improvement. The teachers felt that administrators and even perhaps the NJ DOE were demonstrating a desire to find something wrong rather than having a desire to truly improve teaching and learning.

Isabella discussed the implementation of the Stronge TEPES (2006) and AchieveNJ (2013) coming at the same time as other initiatives in the state, under Governor Christie's administration. She said,

Well I think because it coincided with the Common Core and PARCC, for Language Arts teachers anyway, there was a sense that expectations and standards were rising. And I think that the curriculum that we had reflected that and it was easy to sort of see how you could take that curriculum and make it meet those expectations or work to meet those expectations. So I think for us it created a change.

Kim also spoke to the correlation between the implementation of the Stronge TEPES (2006) and the state standards. She has the perception that teachers were holding themselves more accountable because of the standards that exist in the system. She sees this perception as being positive for instruction and culture, which can be seen when she said,

Personally, I see positives because I think teachers are working a little more diligently and I think teachers are, including myself, making sure that they correlate their lesson plans to the standards. And because I know sometimes when

an evaluation occurs and you're a day or two behind in your lesson plans, you will make sure upfront that the evaluator knows. I think I've seen teachers work harder and more diligently and kind of pay attention to the details more.

Both Isabella and Kim saw a shift in accountability for the teachers; however, there seems to be a direct correlation between this accountability and other factors. The introduction of new standards, as well as a change in the mandated standardized test, created more accountability. Because of these, it was not evident that the Stronge TEPES (2006) alone created this shift.

Barbara perceived this accountability as something that impacted teachers' instruction and planning, but not necessarily all the time. She alluded to teachers being aware of the need to be high performing but also knowing that it did not have to happen all the time. She stated,

I think it definitely made teachers a little bit more aware of how you just have to be on your toes at all times, make sure that you are being an impressive teacher. But actual change-wise I think it's only when they have to. A lot of the teachers know how to make sure that they have things going on when the observation is going on but sometimes don't create those kinds of lessons every day.

This statement indicates that the change was not authentic and lasting, but rather teachers were performing while being observed in order to achieve the score on a summative evaluation.

While each participant spoke about a change occurring because of the implantation and use of the Stronge TEPES, the changes were not always perceived as positive. The quantitative data showed similar results. When participants were asked if

the implementation of the Stronge TEPES (2006) had the ability to create positive change in the district only 24.14% (21) agreed or strongly agreed. There is a strong focus on compliance rather than improvement when teachers and administrators make change to their overall practices. The changes are also being credited to other initiatives, such as the adoption of common curriculum standards and a new assessment system, that were implemented around the same time as the Stronge TEPES. The implementation and adoption of the Stronge TEPES (2006) had the potential to bring about positive changes in the district. Its intended purposes of improved instruction and increased student achievement would have been welcome changes. However, rather than focusing on the possibilities, both administrators and teachers have been focused on compliance so the only changes seen are an increase in paperwork, an increase in accountability, and a decrease in quality instruction.

Administrator Use of Cookie Cutter Feedback

During the interviews, participants were asked questions that led to discussions about their perceptions of the quality of feedback they received on both observations and summative evaluations since using the Stronge TEPES (2006). Many of the participants spoke to the lack of personal connection or narrative that existed in the observations. Additionally, the participants had the perception that observers and administrators were using canned phrases, meaning they were using the same prewritten phrases for multiple teachers. They had the perception that observers were not truly reflecting on the lesson but rather using the performance indicators listed in the rubric as a sort of checklist, which is in direct contrast to the way in which the indicators are meant to be used. The

Stronge TEPES (2006) explains that the performance indicators are a list of examples, not requirements, for each performance standard.

Elizabeth spoke of missing the narratives that teachers used to receive, prior to the implementation of the Stronge TEPES (2006). The district's previous observation tool was completely narrative and allowed the observer to write as much or as little as he/she wanted about anything he/she wanted. There was no rubric and no specific required direction.

I think what I miss the most from it is the personal, kind of touchy-feely thing. Like, in the narratives that we used to get, it was much more, well, it gave you just that warm feeling, like maybe some more expression was used. I feel like now, it's kind of cut and dry, like, the following of standards. I know, in the beginning, I think that was, we felt that a lot, because here, when we switched over, it was the same people who were evaluating us for how many years before, and who were saying all these nice, touchy-feely things. And then all of a sudden, we get more like, "Boom, boom, boom, and boom." Like, "Wait. What happened to the other part?"

The observations and evaluations, as suggested by the Stronge TEPES (2006), were based only on evidence. Observers did not have the latitude to be subjective and draw personal conclusions, leaving the teachers with the feeling of being unappreciated and undervalued.

Both Heidi and Kim spoke to the feedback similarly. Rather than receiving any true feedback or suggestions on how to improve, participants spoke to the observations and evaluations providing a list of events or activities that occurred during the lesson or

year. They did not discuss feedback as a vehicle to improve or make changes to instruction, but rather a very direct account of what is currently happening. Heidi stated,

Most of the time it's true to the sense of observation. They make observations of what's happening. I am very well equipped in that because that's kind of the basis of acting training where you see something and you make observations about what you see and you don't make perceptions about what you're seeing. You're not making a judgment about what you're seeing. I see a lot of listing off of what she does, this is what she does, this is what she does, and this is what she does.

I've felt like it reflected the vocabulary, and the observations do reflect what I do and I'm glad that it's being mirrored in the language. But nothing specific.

Similarly, Kim spoke about the observers listing everything that happened during the lesson. The observations gave a minute by minute list of activities that occurred during the lesson. The perception was positive because the list was accurate; however, she did not speak about observers providing constructive criticism or suggestions on how she could improve upon her instruction. She shared,

Usually it's very detailed, and I'm impressed that they have actually written down every single thing that I've done during that 20 to 30 minutes. Honestly, I haven't seen anything that's totally inaccurate or things that were wrong. Every evaluation has been true to what I did that day, and the year-end observation has been true.

The lack of commentary on suggestions for improvement or acknowledgements of impressive pedagogy should be noted. A list of activities does not necessarily provide

teachers with direction for future instructional decisions, but instead it provides administrators with the ability to give all teachers cookie cutter repetitive feedback.

Some of the other participants expressed a more negative perception regarding the feedback that observations and evaluations were providing them. Francesca had the perception that the feedback was written to fit the form rather than have value for the instruction. She said,

What I am receiving now is the exact same feedback I received the first year. I sometimes think they cut and paste things from one year to the next. It doesn't really matter what I am actually doing because I think it's a lot more cookie cutter, where it kind of fits the particular boxes and for me, it doesn't have any real value or impact on me or my practice.

Gabrielle's perception was noticeably more negative. Her tone indicated frustration with the system, as well as the observers for using it as a compliance tool. She stated,

It is laughable. It is laughable. You know a clown could come in and copy and paste the language that is submitted or encouraged to use in each section. So you know in many cases, it allows us to look at different areas of our progress but it doesn't say anything about it. It doesn't really help us improve our lesson plans.

It's all a performance.

Whether the participants' perception was positive or negative, what they shared during the interviews did not demonstrate that the quality of feedback being given on observations and evaluations was of high quality and based on instructional leadership. Some of the participants seemed to appreciate this because they could continue to do

what they wanted, while others seemed angry that their observations and evaluations had lost any true value.

Cookie cutter feedback, which is being provided to most teachers, does not have the potential to improve or impact pedagogy. If the same feedback is being provided to all teachers, from a canned list, there is no personalization or individualized attention to a teacher's craft. In order to see the usage of the Stronge TEPES (2006) impact instruction in the district, observers must take the time to provide teachers with individualized feedback on strengths, weaknesses, as well as suggestions for future practice.

Instruction Unaffected by the Stronge TEPES

One of the major intentions of AchieveNJ (2013), as well as the Stronge TEPES (2006), was to provide teachers with information and feedback that would guide and improve their instructional practices. However, the qualitative data led to the theme, instruction was unaffected by the use of the Stronge TEPES (2006). This means that teachers were providing the same instruction as prior to the adoption and implementation of the system. Each of the participants was asked to reflect on how the Stronge TEPES (2006) has impacted their instruction. The overwhelming majority, 11 participants, felt that it has had no impact on their instruction at all.

Jennifer, because of the quality of feedback she received, did not feel she had the ability to make any real change to her instruction. She said, "I didn't get the feedback I needed to improve on my practice or to change my practice early enough in order for me to make a change for myself." The inability to make change was based on the timing of the feedback. Despite that AchieveNJ (2013) dictated an increase in the number of

observations, this participant did not perceive the process as providing her with timely feedback.

Adam reflected that perhaps he was the reason that Stronge TEPES (2006) has not had an impact on his instruction. Rather than placing the responsibility for this on either the system or the observer, he said,

No, because I'm kind of set in my ways. And again, I've been a teacher for over 20 years, and I'll listen to suggestions, but some of the feedback is weird. Like, I had a pre-conference and we discussed what I was going to do. I did that. Why wouldn't you say to change it in the pre-conference? So I mean, I understand they have to give you something because otherwise what's the point, but truthfully I don't buy it.

Both teachers and administrators have a responsibility of ensuring that the Stronge TEPES (2006) leads to appropriate changes. If the teachers are unwilling to receive feedback and suggestions, then the system will continue to have no impact on instruction.

Gabrielle felt that she knew when to make changes to instruction, with or without the observations and evaluations. She shared, "It hasn't. I've really just tried to continue to do what I know is best for my students from my experience and my training in spite of this huge albatross." Similarly, Isabella thought that she pushed herself to change, rather than doing it because of the system. She said, "Maybe a little bit, maybe it makes me hold myself to a higher standard in some ways but I think I would have done that anyway." This demonstrates a lack of confidence in the system and portrays it as cumbersome rather than impactful and helpful.

Kim discussed the push to include technology in instruction. However, her response was noncommittal. She said, “Not a whole lot, but yeah I guess you can fix things, and you can concentrate more on certain things. I know technology is a big deal, but it is not my big thing so I ignore those suggestions.” Her response spoke to the culture of teachers disregarding the feedback provided if they do not think it is worthwhile or pertinent.

Several other participants did not elaborate on their responses when asked about the effect on instruction. Mary said, “No”, and Francesca stated, “Not at all.” Their lack of response, similar to Larry’s “none whatsoever” demonstrated a negative tone. It was evident in their demeanor and body language that they were slightly annoyed to have even been asked this question.

However, the participants’ interview responses differed from the quantitative data. When asked if the written feedback had a positive influence on teaching practices, the majority of participants, 47.27% (42), either agreed or strongly agreed. When asked if the written feedback has led to a positive change in teaching practices, the majority of participants, 47.12% (41), either agreed or strongly agreed. The interviews allowed participants an opportunity to reflect more carefully on how effective feedback has been to their instructional practice.

The overall purpose and goal of the Stronge TEPES (2006) is to improve instruction. There were a variety of reasons why the implementation of the Stronge TEPES (2006) did not have an impact on instruction. However, with teachers not making any changes to instruction based on the feedback they are receiving on both observations and evaluations, it can be determined that instruction is neither improving nor declining

in quality. If teachers are making changes to instruction due to other influences (as some participants explained), instruction may be changing, despite the lack of impact the Stronge TEPEs (2006) is having.

Lack of Influence on Professional Development

Prior to the close of each interview, the participants were asked how the Stronge TEPEs (2006) has changed professional development in the district. Many of the teachers did not see a connection between the two. This was evidenced by their discussion of professional development as a completely separate entity from the evaluation system. When responding to the interview questions, many of the participants never mentioned the Stronge TEPEs (2006), and some of them even seemed confused by the implication that there should be a connection between the two. Larry shared, “Not to my knowledge. I don’t think so. There is nothing between the two.” This was a common sentiment shared by most participants.

Some of the participants did explain in some detail the lack of connection or change in professional development and teacher supervision since the implementation of the Stronge TEPEs (2006). Elizabeth stated,

I don't know that it has had any impact. I know that we have the committee (ScIP) that meets, and I know a lot of times, the suggestions that they make for professional development are then overruled by administrators. And that many times over the years, what we want to have as professional development has never happened. So, I don't know what the connection would be between the Stronge and professional development.

The discussion of top down professional development choices for teachers directly contrasts the intended purposes of the School Improvement Panel (ScIP) and the District Advisory Council (DEAC) as outlined by AchieveNJ.

Jennifer did not feel that her observations or summative evaluations affected the professional development she was provided or chose to take. She said,

No. No, not in my personal experience, and though I have only been here for a short time, I don't see any impact on professional development. Although my supervisor may have given me some ideas, it wasn't in observations or related to them. So, not really.

Though she was receiving suggestions on appropriate professional development, Jennifer did not see a correlation between these suggestions and her observations.

Isabella was able to talk about the importance of the professional development opportunities in the district and how she thought they were valuable, but she did not see any evidence of a correlation between the implementation of the Stronge TEPES (2006) and these opportunities.

Let's think about that. I think that a lot of the professional development we had was about backward design and about planning good lessons. So if it's just a question of whether professional development was impacted, I don't think so. I think the professional development was more than that. It was more about best practices and not about observations.

Though the professional development may have impacted professional practices in a positive manner for some individual teachers, it was not because of the implementation or usage of the Stronge TEPES (2006).

When asked directly, as a follow up question, how professional development in the district has been impacted since the implementation of the Stronge TEPES (2006), Heidi replied,

Professional development in the district hasn't changed. I think that my growth is really driven by my own personal quest for deeper content knowledge. That's why I'm going to get my master's in theater studies and not in education, because I feel like if I know my content, if I'm a master of my content, instructional strategies will follow.

The shift in supervision from evaluation to instructional supervision and professional development does not seem evident through the perceptions of the interview participants.

Zepeda (2007) identifies the three main aspects of the supervisory process as instructional supervision, professional development, and evaluation. The implementation and adoption of the Stronge TEPES (2006) has not led to an effective supervisory process as professional development has not changed. There has been no bridge between the identified areas in need of improvement through observations and evaluations and teachers' professional development paths.

Synthesis of Findings

This chapter documented information obtained through the study, the design and the participants. It included data gathered from the online survey, the face-to-face interviews, and the analysis of 36 summative evaluation scores. Throughout the data analysis, several themes consistently emerged, in both the quantitative and the qualitative data: the perception of the implementation, the ability to bring about change, the quality

feedback, no effect on instruction, no impact on professional development and the consistency in scores. Chapter 5 will provide an in depth discussion of the findings.

In both the survey and the interviews, the perception of the implementation of the Stronge TEPES (2006) was addressed by participants. The quantitative and qualitative data did not indicate that the participants had a strong perception, either negatively or positively, about the initial implementation process. There is the appearance of varying understanding and comfort levels with the adoption of the system from the points of view of the participants.

The second theme that emerged was the ability to bring about change. The qualitative data supports two sub-themes: an increase in paperwork and an increase in accountability for instruction. None of the interview responses demonstrated that the participants viewed the change or the sub-themes as positive for the district. The quantitative data on this theme revealed a neutral perception from the participants on whether the implementation and adoption of the Stronge TEPES (2006) had the ability to create positive change in the district.

The next theme that materialized was the quality of feedback provided to teachers from both observations and evaluations. The quantitative data provided the perceptions of the participants on whether the feedback helped to identify strengths in teaching, helped to identify areas needing improvement in teaching, was valuable, was meaningful in identifying teaching practices, led to learning about teaching, had a positive influence on teaching, caused reflection on teaching, and led to positive changes in teaching practices. The qualitative data portrayed the belief that feedback was merely a listing of evidence rather than quality suggestions for improvement.

Following this, the theme of no effect on instruction arose through the data. The qualitative data supports the statement that there has been little to no impact on instruction based on the implementation and adoption of the Stronge TEPES (2006). While a small aspect of the survey addresses this theme, the participants seemed to be more neutral when answering these questions.

Another theme that resulted from the data sources was the lack of no impact on professional development. Both the quantitative and qualitative data demonstrate that teachers do not see a correlation between their observations or evaluations and the professional development or support for improvement in the district.

The final theme was the consistency in scores. One data source was utilized to review this theme. There appears to be a consistency in the scores, not only in buildings but also across the district.

Conclusion

The quantitative and qualitative finds in this study were collected using a convergent parallel design (Creswell & Plano Clark, 2011). Each of the data collection phases had equal value when reaching final conclusions about the implementation, adoption, and usage of the Stronge TEPES (2006) in City Public Schools. Four categories emerged from the quantitative data, and five themes emerged during from the qualitative data.

The chapter that follows will substantially expand upon the findings of the study through summarization and discussion, interpretation, implications, and recommendations and a discussion of the findings. In addition, the final conclusions may

lead to further changes in administrator and teacher practice throughout the district and state.

Chapter 5

Discussion and Implications

Race to the Top (RTT) (2009) was a federal grant made available to states who were willing to use the funding to improve on the teacher evaluation system based on the idea that improved teacher evaluation would lead to improved instruction and in turn greater student achievement. As a way to meet this requirement, the state of New Jersey began to require districts to make changes to then current teacher evaluation practices. Under the AchieveNJ mandates (2013), as part of the TeachNJ regulations (2012), districts had to choose an evaluation model which included a four-point rating scale for a teacher's final evaluation.

This specific policy change introduced the addition of quantitative measures to both observations and summative evaluations that were previously completely qualitatively. It also mandated the inclusion of student achievement in each teacher's summative evaluation. A teacher's summative evaluation is based on a score, between zero and four, and combines teacher practice, student growth objectives, (SGOs), and student growth percentiles (SGPs).

City Public Schools, a pseudonym given to protect the confidentiality of the district, in accordance with the mandates of AchieveNJ (2013) adopted the Stronge TEPES (2006) in 2013. At that time, every teacher in the district received substantial professional development on the requirements of the system before beginning to receive classroom observations and summative evaluations using the Stronge TEPES (2006). In the summer of 2013, the administrators and observers in the district received training on

how to complete both an observation and a summative evaluation ensuring inter-rater reliability.

The Stronge TEPES (2006), one of the NJ DOE approved systems, includes seven standards: professional knowledge, instructional planning, instructional delivery, assessment of/for learning, learning environment, professionalism, and student achievement. The four-point rating scale for this evaluation tool includes: ineffective, partially effective, effective, and highly effective (Stronge, 2006). This mixed methods study examined teacher perceptions of whether the Stronge TEPES (2006) has been successfully implemented in the City Public Schools and if teachers felt they were receiving quality feedback that helped to guide and improve instruction.

This chapter presents a discussion of the findings and themes directly correlated to the research questions that guided this mixed methods study. The data were organized into six themes, each of which was related back to the theoretical framework and existing literature. The implications and recommendations for policy, research, and practice will be discussed, in addition to the possible impacts on leadership. Districts will be able to use these recommendations to inform future practice of administrators, observers and teachers.

Reforms such as the Common Core State Standards (2009), PARCC (2014), and TeachNJ (2012) demonstrate that politicians, educators, community members, parents, and students, would like to see changes to the current K-12 educational system. The changes have affected both the culture and the dynamics of school districts. One of the most drastic of these reforms, especially for teachers, came in the form of new evaluation systems, perhaps because of the lack of buy-in from both teachers and administrators. In

accordance with AchieveNJ (2013), City Public Schools, as well as many other districts, adopted the Stronge TEPES (2006) in September 2013.

Through this mixed methods study, I investigated teacher perceptions of the implementation of the Stronge TEPES (2006) in the City Public Schools, as well as its ability to affect positive change, and the teachers' perceptions regarding the quality of feedback they are receiving on their observations and summative evaluations. The responses of teachers have the potential to inform future practice of administrators, as well as assist in the guidance of professional learning for teachers.

The purpose of this mixed methods study was to examine teachers' perceptions of the Stronge TEPES (2006) with regard to its ability to affect positive change in instruction, student achievement, as well as provide professional growth opportunities for teachers based on individualized feedback in observations and on summative evaluations in City Public Schools.

Through an examination and analysis of the three data sources, a path for a shift in the supervisory model became apparent. The teachers have shared their personal experiences with observation, evaluations, feedback, and implementation through an in-depth analysis of a structured survey, semi structured interviews, and an analysis of individualized summative evaluation scores over a two-year period.

Discussion of Findings

This study gathered data from the online survey, the face-to-face interviews, and the analysis of 36 summative evaluation scores. Throughout the data analysis, several themes consistently emerged, in both the quantitative and the qualitative data: the perception of the implementation of the Stronge TEPES, the potential to bring about

change, instruction unaffected by the Stronge TEPEs, administrator use of cookie cutter feedback, the change in summative evaluation scores, and lack of influence on professional development. These were instrumental in answering each of the research questions.

Perceptions of the implementation of Stronge TEPEs. In both the survey and the interviews, the perception of the implementation of the Stronge TEPEs (2006) was addressed by participants. Neither the quantitative nor qualitative data indicated that the participants had a strong perception, negatively or positively, about the initial implementation process. There is the appearance of varying understandings and comfort levels with the adoption of the new system of evaluation from the point of view of the participants. This has the potential to be a result of the differences in the ways administrators are managing the process throughout the district. The participants in Wacha's (2013) indicated that they were not instrumental in the initial implementation process; much like City Public Schools where they were just participants. While the initial professional development was offered through Central Office administration and a certified Stronge TEPEs (2006) staff member, the follow through varied from building to building. Administrators may have different comfort levels with using the system, and in turn, the teachers in the building were impacted. Administrators with the highest comfort level and understanding were able to provide more extensive and clear professional development.

When asked in both the survey and the interviews about the implementation process, participants referred to the initial professional development and usage of the Stronge TEPEs (2006). Teachers were provided with a comprehensive one day

workshop given by a member of the Stronge TEPES professional development team, as well as follow-up trainings provided by district administrators. The initial implementation and explanation of the system was consistent throughout the district. However, the follow up trainings were building based and at times offered only when necessary. This had an impact on the teachers' perceptions of the implementation process.

Similar to the findings of Wacha (2013), teachers were not involved in the planning of or providing the professional development needed to implement a new evaluation system. The initial survey completed by NJPSA, NJASA, and NJEA (2013) had similar findings. Fifty four percentage of their participants did not believe the training on a teacher evaluation system in their districts was adequate. Contrasting, 74% of the administrators and superintendents viewed the training on a teacher model as adequate (NJPSA, NJASA, & NJEA, 2013). This lack of collaboration between the administrators and the teachers furthered the decrease in staff morale and the belief by teachers that the Stronge TEPES (2006) was being used as a punitive measure and not a tool to enhance teacher learning and instruction.

While the data demonstrates neither a strong perception in one direction nor the other by the participants, it can be concluded that the participants felt that the implementation was successful and provided teachers and users with the knowledge necessary to effectively use the Stronge TEPES (2006). In contrast to Towe (2012) who found that participants viewed the formative process as having limited impact on improved teacher practice. This conclusion can be made because in the survey the participants were asked about the overall quality of the implementation and the majority

of responses were 3 or higher. In addition, the interviews resulted in positive data for the implementation process.

However, it can also be inferred that usage of the system, as seen by the participants, is drastically different from the system providing an avenue for improvement. The Stronge TEPES (2006) provided a way in which to implement the changes from AchieveNJ (2013) while allowing district personnel the ability to be autonomous. The data does not support that the implementation of the Stronge TEPES (2006) in City Public Schools, while implemented effectively, as most participants indicated, accomplishes the overarching goals of the NJDOE's mandates in City Public Schools.

The potential to bring about change. The second theme that emerged was feedback received from the Stronge TEPES's ability to bring about change in the district. The qualitative data supports two sub-themes: an increase in paperwork and an increase in accountability for instruction. None of the interview responses demonstrated that the participants viewed the change or the sub-themes as positive for the district. Towe (2012) came to a different conclusion; Towe's (2012) participants indicated that the teacher evaluation process had some degree of influence on improving their teaching. Some responses spoke to specific pedagogical skills like differentiated. In this study, the quantitative data on this theme revealed a neutral perception from the participants of whether the implementation and adoption of the Stronge TEPES (2006) feedback had the ability to create positive change in the district.

The quantitative data, which spoke directly to increased student learning, one of the intended changes from AchieveNJ, demonstrated that the teachers in City Public

Schools saw no evidence of a connection between the implementation of the Stronge TEPES (2006) and improved student achievement. As previously noted, the two primary goals of AchieveNJ (2013) were to improve instruction and to improve student achievement throughout the state. The mandate allowed districts the choice as to which system to use to accomplish these goals. The data from this study indicates that the choice of the Stronge TEPES (2006) in City Public Schools has not accomplished one of the goals, increased student achievement. When making significant changes, such as a change to a teacher evaluation system, there should be buy in and belief in the system by the teachers. Only with this buy in will teachers internalize the need for this change and allow it to impact their instruction. In order to improve student achievement, instruction needed to change. Teachers have not changed or shifted their instruction because of the implementation of the Stronge TEPES (2006). This is because they did not initially see the need for change, did not buy in to the change, and have not received feedback that resulted in changes to instruction. Callahan and Sadeghi (2015) when presenting their findings saw no change in the way teachers taught their students, the emphasis they placed on raising standardized test scores, or teachers' knowledge and understanding subject area.

Instruction unaffected by the Stronge TEPES. The Stronge TEPES (2006) requires the observer to collect a preponderance of evidence over a school year in order to provide teachers with a summative evaluation score. For the purpose of this research, the evidence collected includes observations, both formal and informal, the documentation log, student surveys, SGOs, and SGPs, if applicable. Both teachers and administrators are responsible for various components and data collection. The

administrators are responsible for the completion of the observations and the summative evaluations, as well as approving the SGOs written by teachers. However, the teachers are responsible to complete a documentation log, to distribute a student survey, to analyze the results of the student surveys, and to write and collect data for SGOs. According to the survey completed by NJPSA, NJASA, and NJEA (2013), teachers did not feel that the creation of SGOs was a collaborative process. The NJ DOE is responsible for the release of SGP scores.

When asked about the opportunity for the Stronge TEPES (2006) to create change in the district, participants discussed an increase in paperwork for compliance. This is a concern because if both teachers and administrators are forced to spend time completing paperwork that has little to no meaning they do not have ample time to spend planning quality instruction and affecting change in a building. They did not indicate that this increase of data collection to get a summative evaluation and score affected their instruction in a positive manner. Similar to the findings of Curran (2014) when studying the teacher perceptions of the Texas PDAS, teachers felt that the inclusion of the evaluation systems created time constraints for the administrators leading the less valuable and useful feedback.

Additionally, the teachers who were interviewed spoke to the perception of an increase in instructional accountability. While this initially sounded like a positive shift based on implementation, when further clarifying, participants indicated that they were providing a higher level of instruction when being observed in order to receive a higher rating on their summative evaluation. Rather than being intrinsically motivated to be an effective teacher, this shift occurred for compliance on a need basis. It was not a habitual

change. This is a result of the teachers' desire to receive a high rating on their summative evaluation. There is an emphasis placed on the score received rather than the meaning or feedback that led to the score. Towe (2012) found in her study that participants felt there was little to no true impact on instruction as a result of the implementation of a teacher evaluation system.

Both the qualitative and quantitative data shows that the changes which occurred in the district following the implementation of the Stronge TEPES (2006) did not positively influence student achievement or teacher practice. Shepard (2013) only found that 20% of participants thought the evaluation process had a strong impact on instruction. Time was taken away from the teachers' opportunity to plan and deliver quality instruction due to an increase in paperwork; the teachers have not seen an increase in student achievement since the adoption; and the only positive changes to instruction occur to observations rather than on a routine basis.

Administrator use of cookie cutter feedback. This research sought to find out how teachers perceive the quality of feedback they are receiving in observations and evaluations using the Stronge Teacher Effectiveness Performance Evaluation System.

The quantitative data identified the perceptions of the participants regarding the feedback helped to identify strengths in teaching, helped to identify areas needing improvement in teaching, was valuable, was meaningful in identifying teaching practices, led to learning about teaching, had a positive influence on teaching, encouraged participants to reflect on teaching, and led to positive changes in teaching practices. The responses from participants are a direct reflection of the work of the administrators in the district and the teachers' willingness to accept the feedback. Their perceptions come

from the feedback they are receiving from the various administrators, as well as personal reflection on choices they have made. While teachers felt that that feedback identified strengths and weaknesses in instruction, it did not necessarily lead to changes in instruction or true reflective practices. Sheppard (2013), after conducting a study in rural Georgia, found that the feedback being provided to teachers was having an impact on instruction. The difference between Shepard's (2013) findings and this study can be attributed to a variety of factors including the size of the district, the location of the district, and perhaps the intensity and demands of the difference evaluation systems.

The qualitative data portrayed the belief that feedback was merely a listing of evidence rather than quality suggestions for improvement. This list of evidence does not inform teacher practice for the future. The participants also indicated that the feedback received on both observations and summative evaluations were copied from the list of performance indicators provided in the rubrics of the Stronge TEPES (2006) and not actually related to their specific lesson or instruction. Curran (2014) found that the perception was that teachers were receiving less feedback during observations and evaluations. Similar to this study, teachers being observed using a formal evaluation tool are not receiving either enough or high quality feedback that can truly make an impact on their instruction.

The data, both qualitative and quantitative, led to the conclusion that administrator' practice, similar to teacher practice, has not been positively influenced. The capacity of administrators to complete the work outlined in AchieveNJ (2013) has been a concern for all educator groups in the state since its adoption (NJPSA, NJASA, & NJEA, 2013). As perceived by the teachers, administrators have seen an increase of

paperwork, which has led to insufficient and low quality feedback because of time parameters. Administrators' practice is focused on compliance with both the requirements of AchieveNJ (2013), as well as the requirements to satisfy completion of the Stronge TEPES (2006) each year. Administrators are given completion dates by both the regulations and the district. These dates place time parameters on an administrator's ability to provide quality feedback to teachers. Wacha's (2013) findings indicate that the feedback received was timely, just not high quality. In addition, the increase in the number of observations has led to less time for administrators to spend on their other responsibilities and duties. They cannot continue to visit classrooms regularly or be instructional leaders in the buildings.

The change in summative evaluation scores. Two consecutive years of summative evaluation teaching scores of teachers in the City Public Schools were analyzed to determine if there was consistency in the use of the Stronge Teacher Effectiveness Performance Evaluation System throughout the district.

One data source was utilized to review this theme. There appears to be consistency in the scores, not only in buildings but also across the district. After analyzing the scores, it became evident that the scores from one year to the next, for the majority of the teachers, did not have a drastic change, if any change at all. This either indicates that teacher practice has not been influenced from year to year or that administrator practice has not been influenced from year to year. It is possible that the scores have not changed because teacher practice is the same; observers are seeing similar instructional methods and professional decision making from one year to the next. It is also possible that scores have not drastically changed because administrators are

using the previous year's summative evaluations and scores as a model for the current year.

Lack of influence on professional development. In order to determine if the Stronge TEPES (2006) can be used to shift the supervisory model in City Public Schools from evaluation to instructional supervision and professional development (Zepeda, 2007) based on the perceptions of teachers, both qualitative and the quantitative data were collected. Both indicate that there has been no impact on professional development since the implementation and adoption of the Stronge TEPES (2006). Both the quantitative and qualitative data demonstrate that teachers do not see a correlation between their observations or evaluations and the professional development or support for improvement in the district. Wacha (2013) had a similar finding; the participants saw little to no correlation between the evaluation system and the professional development provided to teachers from the district. Callahan and Sadeghi (2015) found that less than 40% of their participants saw a correlation between AchieveNJ evaluations and individual professional development plans. It is evident that districts and schools should begin to move through the levels of teacher supervision (instructional supervision, professional development, and evaluation) as Zepeda (2007) defines. Currently, we are at the level of evaluation but have not used the evaluation as a catalyst for instructional leadership or professional development

AchieveNJ (2013) requires that each school has a ScIP. The purpose of this team of educators, both teachers and administrators, is to inform professional development decisions from teachers' observations and evaluations. The ScIP is supposed to identify weaknesses and provide teachers with an opportunity to learn and

improve. It can be concluded that the SciPs in City Public Schools either do not have access to teachers' observations and evaluations or are not properly using the data to inform decision making as participants were unable to even identify a correlation between the two. The supervisory model has not changed from evaluation to instructional supervision. Based on this, it can also be concluded that the administrators in the district are not viewed as transformational or instructional leaders.

Research Questions

In this study, I attempted to value of the implementation of the Stronge TEPES (2006) in City Public Schools, the quality of feedback received by teachers on classroom observations, and the changes to instruction based on quality professional growth opportunities. Through this study, I hoped to determine if teachers in the district have experienced positive change in their instruction based on the implementation of the Stronge TEPES (2006). The verification of my findings provided an answer to each of the research questions.

Research question one. The first research question sought to determine how teachers in the City Public Schools perceived the implementation, specifically the ability to create positive change, of the Stronge TEPES (2006). Both qualitative and quantitative data were collected to answer this question. Despite participants' perceptions of the implementation of the Stronge TEPES (2006), the data indicates that the implementation was successful and provided the teachers throughout the district the information needed to use the program. However, the evidence of change to instructional practice, as seen in the data, is not positive. There was no evidence of an increase in

student learning or student achievement as a result of implementing the new evaluation system. In addition, there is no evidence of lasting shifts in instruction.

Research question two. The second research question attempted to find out how teachers in the City Public Schools perceived the quality of feedback they were receiving in observations and evaluations using the Stronge TEPES (2006). This question speaks directly to the practices by the administrators in the district. The feedback, while identifying both strengths and weaknesses in instruction, was most often merely a listing of the performance indicators on the Stronge TEPES (2006) rubrics rather than suggestions on how instruction could be impacted. Participants perceived the feedback to be both cookie cutter and generic rather than lesson specific and helpful. It can be concluded that the feedback is being provided as a means of compliance and completion rather than an actual vehicle for improving instruction and student achievement.

Research question three. The third research question sought to determine if the summative evaluation scores of teachers in the City Public Schools demonstrated consistency in the use of the Stronge TEPES (2006). Through the analysis of the quantitative data source on this topic, it is evident that there is consistency in the summative evaluation scores in the district. This demonstrates that more often than not the administrators in the district are using the Stronge TEPES (2006) similarly. However, it can also be concluded that this consistency has not led to growth or improvement for teachers. They are often receiving similar summative evaluation scores from year to year, which indicates that instruction is not being impacted.

Research question four. The final research question provided answers on how the Stronge TEPES (2006) can be used to shift the supervisory model in City Public

Schools from evaluation to instructional supervision and professional development. The evidence reflects the belief that participants saw no connection between evaluations and professional development in the district. Administrators must begin to change their practices and make a true correlation from the classroom observations to the summative evaluations to the professional development plans for teachers.

Reflection on Theoretical Framework

This study was framed around Zepeda's (2017) supervision model of instructional supervision, professional development, and evaluation. Zepeda (2017) argues instructional supervision, professional development, teacher evaluation, and other reform efforts must form a "seamless web" (pg. 8) that when woven together contribute to the overall improvement of teaching and learning and increased student achievement. Zepeda (2007) identifies the three main aspects of the supervisory process as instructional supervision, professional development, and evaluation, and asserts it is within the power of leadership to shape or reshape culture to stimulate teachers to improve instruction (Zepeda, 2017). By creating a collaborative culture, as is the intent of City Public Schools, teachers and leaders are more willing to take risks and learn from their missteps (Zepeda, 2017). Burant (2009) suggests that instructional supervision and professional development are dependent on one another. This study adds that professional development and instructional supervision can also be linked with evaluation as Zepeda (2007) states.

The Stronge TEPES (2006), implemented in the City Public Schools in 2013, is being used at the evaluation stage of the supervisory process. While the original intentions of both AchieveNJ (2013) and the Stronge TEPES (2006) were to improve

instruction and in turn increase student achievement, this study demonstrates that there is a lack of instructional supervision and no emphasis on professional development, the other two aspects of Zepeda's (2007) supervisory model. For the Stronge TEPES (2006) to have the ability to truly create change in the district, both administrator and teacher practice need to shift to a collaborative culture focused on improved instruction and increased student achievement (Zepeda, 2017). Administrators must provide teachers with quality feedback directly connected to future professional development and growth for teachers, and teachers must be willing to accept and internalize the feedback.

Similarly, Chen and McCray (2012) discuss the whole teacher approach to teacher supervision inclusive of professional development. Much like looking at the whole student, the whole teacher approach looks to support the teacher in multiple ways. Both Zepeda (2007; 2017) and Chen and McCray (2012) emphasize the importance of differentiating between evaluation and supervision of teachers. Zepeda (2017) defines this as that which "aims to promote growth, development, interaction, fault free problem solving, and a commitment to build capacity in teachers" (pg. 29). Supervision of teachers is more than just observing and evaluating them; it is focused on helping them to improve through professional development. The findings from this study demonstrate that the perception of the teachers in City Public Schools is that the supervision of teachers in the district is focused on observing and evaluating them (perhaps for compliance purposes), rather than working collaboratively to problem solve (Zepeda, 2017; Chen & McCray, 2012). Participants in this study did not see any focus in the observations and evaluations on helping them to improve through professional development. The administrators should shift their goals and purposes of the evaluation

system to instructional supervision and professional development allowing administrators to focus their supervision on a more positive approach rather than one that is punitive.

Zepeda (2017) establishes a clear connection between instructional supervision of teachers and professional development through her framework. Data obtained through observations and evaluations should be used to inform the planning and development of professional growth opportunities for staff (McQuarrie & Woods, 1991; Zepeda, 2017). In addition, Zepeda (2017) discusses other approaches to teacher supervision, such as clinical supervision, peer coaching, cognitive coaching, and mentoring. Each of these approaches can result in using data further to inform instructional decision making, generate professional growth opportunity for teachers, and strengthen instructional supervision.

The purpose of teacher evaluation in the state of New Jersey which models the recommendations of Zepeda (2017) and Chen and McCray (2012), inclusive of both TeachNJ (2012) regulations and the AchieveNJ (2013) mandates, is to improve teacher practice and student achievement. Unfortunately, participants in this study indicated that they saw no long term improvements in either teacher practice or student achievement. They felt teachers were able to demonstrate quality pedagogy, but because of the lack of instructional supervision and professional development, it only occurred as a result of their desire to score higher on a summative evaluation rather than improve quality of instruction and student achievement.

The focus of improving teacher practice and providing teachers with an opportunity to access quality professional development should be at the forefront of teacher observations and summative evaluations. This will bring positive change focused

on adult learning and teacher development. The participants in the study demonstrated that the feedback they are receiving, as well as their personal willingness to change their instruction, has not created the needed shift in culture to move away from punitive measures based on teacher evaluation to ones of learning and growth for educators.

Leadership

District and school administrators can no longer be merely managers of buildings, or transactional leaders (Shields, 2013) and while the day to day activities often threaten to steer administrators from their true work, it is essential that they maintain a laser focus on transformational leadership (Shields, 2010). The primary purpose of schools is to educate children through quality instruction. Therefore, the primary purpose of school level and district level leaders is to provide an opportunity for this education to occur by developing quality teachers and instructional leaders. The Stronge TEPES (2006), TeachNJ (2012), and AchieveNJ (2013) reflect the same purposes: the improvement of instruction and an increase in student achievement. However, if the leaders of districts and buildings are acting as transactional leaders (Shields, 2010) rather than having a true focus on teaching and learning, change is needed.

In a mixed methods study to determine how leadership impacted school improvement, Klar and Brewer (2013) discussed four core categories for whole school or comprehensive school reform: setting directions, developing people, redesigning the organization, and managing the instructional program.

Developing people refers to providing individualized support and consideration, offering intellectual stimulation, and modeling appropriate values and practices (Klar & Brewer, 2013). Likewise, the requirements for observations and evaluations through

AchieveNJ (2013) are an opportunity for the administrators in City Public Schools, as well as other districts, to develop educators as defined by Klar and Brewer (2013).

During observations, leadership must identify teacher deficiencies and share these during the post-conference, which provides a time to develop a plan for professional development and growth on an individual basis.

In order to change an educational institution for the better, instruction needs to improve. While this is not the only means to achieving academic success, it is essential to establishing a culture of high expectations and improvement. Rather than only focusing on the completion of the observation and summative evaluations as a means of compliance, leaders must focus on the conversations with teachers and the professional development that follows. All of this happens through high quality feedback to teachers and then professional development opportunities that will assist them in being more reflective in their practices.

Recommendations

Based on the findings of this study, it is recommended that City Public Schools, and other districts having similar experiences, take steps to improve the practice and policies existing in the district. The recommendations are correlated to each of the research questions that guided this study.

First, it is recommended that the district provide professional development for all teachers and administrators to review the Stronge TEPES (2006) and AchieveNJ (2013). The purposes and goals of each should be considered when planning for this. Since time has passed since the initial implementation, this will provide an opportunity to reinvigorate the usage of the system. In addition, it is recommended that the professional

learning communities (PLCs) in the district begin to use observation and evaluation data for action research.

It is also recommended that administrators in the district receive extensive professional development on the usage of the Stronge TEPES (2006). Rather than solely focusing on inter-rater reliability, administrators should reexamine the feedback they are writing on observations and evaluations. In addition, administrators should take steps to correlate the feedback provided to teachers directly to the professional development planned for the district. One should drive and inform the other; these should not be separate entities.

Implications

This research is critically important based on the current state of education in New Jersey. The morale of teachers has declined because of the perception that teacher evaluation is punitive or from the lack of consistency and value in observations (NJPSA, NJASA, & NJEA, 2013). However, it is essential that we redirect our work in the area of teacher evaluation and supervision. By identifying the perceptions that currently exist, we, as educators, can reevaluate our evaluation practices and instead of completing classroom observations and summative evaluations as a way to be compliant with TeachNJ (2012) and AchieveNJ (2013), administrators and teachers can use observation data as a basis for professional growth and improvement of teacher practice and in turn, increase student achievement by improving instruction.

Practice. This mixed methods study has the potential to influence both teacher and administrator practice in the future. Teacher evaluation, as a result of AchieveNJ (2013), has created a major shift in the responsibilities of administrators. It has increased

the number of times teachers are formally observed, and in turn, it has also increased the amount of observations that administrators are responsible for completing, decreasing the amount of time spent in other areas.

In order to redirect our practice from a focus on compliance with the regulations, both administrators and teachers can create professional growth plans resulting in both improved instruction and increased student achievement, two of the primary goals of AchieveNJ (2013). It is recommended that administrators take the time they are spending in classrooms as an opportunity to truly influence instruction. By providing teachers with high quality feedback, regarding the strengths and weaknesses of instruction, as well as suggestions for improvement, administrators can become change agents.

Teachers can also begin to take on a different point of view of the Stronge TEPES (2006). Because it was adopted during a time when teachers were feeling vulnerable in the eyes of the public and the state politicians, the implementation and adoption was tainted with a negative perception. As educators, we cannot combat the ever changing politics that surround the state and at times influence decisions for our profession. However, we can be cognizant of the impact they have on our actual practice. By being more self-reflective and open to feedback from administrators, teachers can adopt new instructional practices. By viewing the documentation log and the student surveys as an opportunity to be grow, teachers can change their practice from compliance to self – reflection and growth.

Policy. TeachNJ (2012), as well as the accompanying regulations of AchieveNJ (2013), were the leading catalyst for the adoption of the Stronge TEPES (2006) in City Public Schools. As the NJ DOE continues to examine the policy, changes are being

made. For example, when AchieveNJ (2013) was implemented in September 2013, tenured teachers were required to be observed a minimum of three times during each school year. However, in September 2016, AchieveNJ (2013) changed the regulations to only mandate two observations of tenured teachers during a school year based on the feedback of the state's educators and administrators.

As the political landscape in the state has once again changed, various groups are proposing additional changes because now the educational and political leaders would like to shift the balance back in favor of the teachers and professional growth (NJSBA, 2019). For example, the NJ DOE is working with NJPSA to promote the Connected Action Roadmap (CAR) to promote many of the same reforms introduced six years ago, including a standards based curriculum, assessment, and professional development (NJSBA, 2019). The introduction of new standards, new standardized tests, and a new teacher evaluation system all under Governor Christie is being said to have created both initiative fatigue and poor morale. While I do not think education should return to the days of completely subjective observation tools, I do believe this study demonstrates a need for further policy change. The increase in paperwork, as well as the multitude of changes occurring at the same time, has left many confused and unmotivated. We have to create a direct correlation between teacher observation and evaluation and professional development. The adoption of CAR (NJSBA, 2019) throughout the state may provide an opportunity for both teachers and administrators to refocus their efforts. By making changes that allow administrators the time to adequately provide feedback on instruction, as well as relate it to suggestions for professional growth, we can move along the continuum of Zepeda's (2007) supervisory model.

In addition, AchieveNJ (2013) does not have specific requirements for the observations and summative evaluations of supervisors, directors, and various other educational personnel. Because of this, districts have shifted their administrators in to roles that are not covered by the TeachNJ (2012) regulations and AchieveNJ (2013) mandates. Titles have been changed in order for administrators to be excluded from the mandates and regulations. However, this will not improve education. This will only keep education caged (Hess, 2013). Not only do leadership need to be honest about the performance of supervisors, directors, and similar personnel, but I also think we need to use data like problem solvers (Hess, 2013). We may eventually see the inclusion of all certified staff members in both TeachNJ (2012) and AchieveNJ (2013), rather than only teachers, principals, and assistant principals.

Future Research

Further research should be conducted to determine if the implementation and adoption of the Stronge TEPES (2006) in other districts throughout the state has led to similar outcomes: lack of quality feedback with a focus on compliance. This study was conducted in one setting; however, the system was adopted in over 100 districts across the state. It would be beneficial to use the same instruments in other settings with other participants prior to making policy changes at the state level.

In addition, future research should be conducted on the implementation of AchieveNJ (2013) and TeachNJ (2012). These mandates and regulations were implemented and adopted in the state of New Jersey during a time when political forces were creating a negative view of teacher practice. Because of this, the impacts of AchieveNJ (2013) and TeachNJ (2012) should be researched. The purpose of both was

to improve instruction and student achievement. This can be measured to ensure that the mandates and regulations are accomplishing what they were intended to accomplish, rather than create punitive measures for the teaching profession.

Finally, there is a need through research to continue to establish the connection between observations and evaluations and professional growth opportunities, as McQuarrie & Woods (1991) discuss. The supervision of teachers should continue to be researched in order to ensure that we are meeting the individual needs of staff in order to see professional growth.

Limitations

One limitation that impacted this study was the changes that the NJ DOE made to AchieveNJ (2013) during the course of this study. When the Stronge TEPES (2006) was implemented in City Public Schools, it was required that tenured teachers were observed three times a year. One observation needed to be the full length of a lesson, while the other two were only 20 minutes. Since then, AchieveNJ (2013) changed its requirements to only two observations per year for tenured teachers. Additionally, the time limits were completely removed and the weighting scale for the components of the summative evaluation (teacher practice, SGOs, and SGPs) were changed several times. The most recent change was in September 2018. This changed the impact that student performance and achievement has on a teacher's summative evaluation score and employment.

A second possible limitation was the change in staff that occurred within the district. However, since all of the data were collected during the same school year, it only effected the summative evaluation score analysis. By using criterion sampling

(Patton, 2002) followed by simple random sampling (Fowler, 2009), this limitation did not impact the study. The criterion used was that only teachers with two consecutive years of summative evaluations scores could be considered. From this group, simple random sampling was used to select six teachers in each building.

An additional potential limitation to the study was the willingness of the teachers to be participants in the study. The completion of the survey, as well as the participation in the interviews, were all based on the willingness of teachers. The data was reliable because there was a 28% return rate on the interviews, and 23 teachers were willing to be interviewed, though only 13 interviews were conducted due to data saturation.

A final limitation to the study was my position as an administrator and observer in the district. I was concerned that teachers would be unwilling to share honest feedback during interviews because of this. However, it was evident through their willingness to share negative opinions and experiences that this may have limited findings.

Trustworthiness

The trustworthiness of a research study is important to evaluating its worth (Lincoln & Guba, 1985). Trustworthiness involves establishing confirmability, credibility, dependability, and transferability. Without trustworthiness in research, it will make the study, as well as the findings, conclusions, and recommendations less powerful. It ensures that the research is reliable and valid.

Confirmability. Confirmability is the degree of neutrality or the extent to which the findings of a study are shaped by the respondents and not researcher bias, motivation,

or interest (Lincoln & Guba, 1985). Because it is my belief that teacher practice has not been positively impacted by the implementation of new evaluation systems, confirmability was a concern. Through this mixed methods study, I investigated teacher perceptions of whether the Stronge TEPES (2006) has been successfully implemented and if teachers feel they were receiving quality feedback that can help guide and improve instruction.

All of the findings, recommendations, and conclusions are based on the responses of the participants rather than beliefs and biases of the researcher. The survey answers, the codes from the analysis of the interviews, and the analysis of the summative evaluation scores were used to inform the answers to the research questions.

Credibility. Lincoln and Guba (1985) describe credibility as the confidence that exists in the truth of the findings. In this study, credibility will come through the triangulation of the data. Creswell (2014) discusses triangulation as a means to use more than one method to collect data on the same topic. This is a way of assuring the credibility of research through the use of a variety of methods to collect data on the same topic, which involves different types of samples as well as methods of data collection. Through triangulation, I corroborate my findings, as well as facilitated a deeper understanding. This mixed methods study included three sources of data: survey data, interview data, and summative evaluation data. This variety of methods to collect data on the same topic created credibility in the study. Each of the topics in this research study: the implementation, the ability to affect change, the supervisions of teachers, and the quality of feedback, had data from more than one source creating credibility and triangulation.

Dependability. When looking at this research study for dependability, showing that the findings are consistent and could be repeated (Lincoln & Guba, 1985), it is evident that the study has dependability because it can be repeated in a different setting. As a result of Race to the Top (2009), other states in the country are seeing changes to their evaluation system as evidenced through Sheppard (2012) in Georgia who studied the Teacher Evaluation Profile. As previously stated, the Stronge TEPEs (2006) is used in other states such as Arizona, Florida, Georgia, Kentucky, Texas, Virginia, and Wisconsin (Stronge, 2015). School districts in any of these states or even another district in New Jersey could use a similar instrument to determine teacher perceptions about the implementation process, the ability to create positive change through the system, the quality of feedback, and the potential to steer teachers' professional growth.

Transferability. Lincoln and Guba (1985) describe transferability as whether this study could be applicable in other contexts. Transferability exists in this study in two ways, in comparison to other evaluation tools and to other districts. While I surveyed the teachers from the City Public Schools, this study has transferability with other evaluation systems and perhaps even other districts using the same system. It is possible to use a similar instrument to survey educators about either the Danielson (2007) or the Marzano (2013) framework. Additionally, the Stronge TEPEs (2006) is used in states other than New Jersey. Because of this, the study could be repeated, using a similar instrument, in these other states. Both of these give the study transferability.

Conclusion

This study did not look at the supervisory model in the same way that Minnear-Peplinski (2009) did, but it does reiterate the need for a change in the use of the Stronge

TEPES (2006), as well as other evaluation tools. As this study examined the perceptions of teachers, it demonstrated that observations and evaluations need to lead toward professional growth rather than punitive measures. In the future, the feedback teachers receive should be valuable and lead to professional growth, improved instruction, and increased student achievement. Both observations and evaluations can become a map and path to professional development and improved instructional supervision, two aspects of the supervisory model discussed by Zepeda (2007).

When examining the current literature, only two of the studies were conducted in the state of New Jersey, and Towe (2012) focused her work on the Danielson Framework (2007), rather than the Stronge TEPES (2006), the focus of this study. The studies conducted outside of the state draw conclusions about state specific evaluation tools. This study addressed the gap in the literature regarding the Stronge TEPES (2006). It focused on the teachers' perceptions of the implementation, the quality of feedback, the consistency in the summative evaluation scores, and the focus on professional growth.

When examining the current literature, there was a lack of information on the Stronge TEPES (2006), as well as AchieveNJ (2013). The Stronge TEPES (2006) can be considered relatively new, as compared to the Danielson Framework. Because the Danielson Framework originated as a teaching framework rather than an evaluation tool and was created in the mid-1990s, districts and educators were familiar with this work prior to AchieveNJ (2013) and many districts used their prior knowledge of the framework as a leading reason to choose it as an evaluation tool.

By shifting the focus of our efforts in the area of teacher evaluation from the current penal system of evaluation, which was created by the previous Governor, to a

system that helps teachers grow professionally, the Stronge TEPES (2006), as well as the other evaluation systems, has the potential to create positive change in education for administrators, teachers, and students resulting in improved instruction and increased student achievement.

References

- AchieveNJ (2013) website. (2013). <http://www.nj.gov/education/AchieveNJ/>
- Achieve the Core Website. (n.d.). Retrieved from <http://achievethecore.org/>
- Blackburn, B. (April 6, 2011). New Jersey Governor Chris Christie Calls His State's Teachers Union 'Political Thugs'. Retrieved from: <http://abcnews.go.com/Politics/jersey-governor-chrischristie-calls-teachers-union-political/story?id=13310446>
- Bouie, Jr., Ed.D., E. L. (2012). The impact of bureaucratic structure, scientific management, and institutionalism on standards-based educational reform. *Mercer Journal of Educational Leadership*, 1(1), 1-21. Retrieved from <https://libraries.mercer.edu/ursa/bitstream/handle/10898/351/Impact%20of%20Bureaucratic%20Structure,%20Scientific%20Management,%20and%20Institutionalism%20on%20Standards-Based%20Educational%20Reform.pdf?sequence=3>
- Burant, S. E. (2009). The relationship of instructional supervision and professional learning communities as catalysts for authentic professional growth: A study of one school division (Master's thesis). Saskatoon: University of Saskatchewan. http://library2.usask.ca/theses/available/etd-09202009165813/unrestricted/Burant_Thesis_FINAL_August_2009.pdf
- Callahan, K. & Sadeghi, L. (2015). Teacher Perceptions of the Value of Teacher Evaluations: New Jersey's AchieveNJ. Retrieved from <https://pdfs.semanticscholar.org/d6f3/2a0778f9a9ba517521d5e9a38a74fa0c166a.pdf>
- Chen, J., & McCray, J. (2012). Conceptual framework for teacher professional development: The whole teacher approach. *NHSA Dialog: A Research-to-Practice Journal for the Early Childhood Field*, 15(1), 8-23. <http://dx.doi.org/10.1080/15240754.2011.636491>
- Chuong, C. (2014). The inconsistent implementation of teacher evaluation reforms. Retrieved from <http://educationnext.org/inconsistent-implementation-teacher-evaluation-reforms/>
- Common Core State Standards Initiative website. (2014). <http://www.corestandards.org>
- Creswell, John. (2014). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. Thousand Oaks, CA: Sage.
- Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research* (2nd ed.). Thousand Oaks, CA: Sage.

- Curran, C. L. (2014). *In-service teacher perception of feedback from formative evaluation within the teacher appraisal process and its relationship to teacher self-efficacy* (Doctoral dissertation). Retrieved from <http://search.proquest.com.ezproxy.rowan.edu/pqdtglobal/docview/1719539263/5A7F951E2E24FF6PQ/1?accountid=13605>
- Danielson, C. (2007). *Enhancing professional practice: A framework for teaching* (2nd ed.). Alexandria, VA: Association for Supervision and Curriculum Development.
- Danielson, C., & McGreal, T. (2000). *Teacher evaluation to enhance professional practice*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Denzin, NK. (1978). *Sociological Methods*. New York: McGraw-Hill.
- Doerr, S. E. (2012). Charlotte Danielson's theory of teacher evaluations: A quantitative study of teachers' perceptions on the four domains. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3499681).
- Doherty, K., & Jacobs, S. (2015, November 2015). State of the states: Evaluating teaching, leading, and learning. *The National Council on Teacher Quality*, 1-160. Retrieved from <http://www.nctq.org/dmsView/StateofStates2015>
- Firestone, W. A., Blitz, C. L., Gitomer, D. H., Kirova, D., Shcherbakov, A., & Nordon, T. L. (2013). *New Jersey teacher evaluation, RU-GSE external assessment, year 1 report*. Retrieved from <http://www.state.nj.us/education/archive/EE4NJ/presources/RUGSE11-12.pdf>
- Fleisher, L. (2010, September 28). Gov. Christie pushes for teacher merit pay based on students' performance. Retrieved from: http://www.nj.com/news/index.ssf/2010/09/gov_christie_advocates_merit_p.html
- Fowler, F.J. (2009). *Survey research methods* (4th ed.). Thousand Oaks, CA: Sage.
- Gardner, D. P. (1983). *A nation at risk*. Washington, DC: The National Commission on Excellence in Education, US Department of Education.
- Gibbs, G. (2007). *Analyzing Qualitative Data*. Thousand Oaks, CA: Sage.
- Hess, F.M. (2013). *Cage-Busting Leadership*. Cambridge, MA: Harvard Education Press.
- Hunter, R. (Ed.). (2004). *Updated edition Madeline Hunter's mastery teaching: Increasing instructional effectiveness in elementary and secondary schools*. Thousand Oaks, CA: Corwin Press

- Johnson BR, Onwuegbuzie AJ. Mixed methods research: a research paradigm whose time has come. *Educational Researcher*. 2004;33 (7):14–26. doi: 10.3102/0013189X033007014.
- Klar, H.W. & Brewer, C.A. (2013). Successful leadership in high-needs schools: An examination of core leadership practices enacted in challenging contexts. *Educational Administration Quarterly*, 49(5), 768-808. doi:10.1177/0013161X13482577
- Kurabinski, M. J. (2010). English language arts. Retrieved from <http://www.state.nj.us/education/aps/cccs/lal/>
- Lincoln, Y., & Guba, E. (1985). *Naturalistic inquiry*. Newbury Park, CA: Sage.
- Manna, P. (2006). Control, persuasion and educational accountability: Implementing the no child left behind act. *Educational Policy*, 20(3), 471-494. doi: 10.1177/0895904805284055.
- Marzano, R. J., Frontier, T., & Livingston, D. (2011). *Effective supervision: Supporting the art and science of teaching*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Marzano, R. J., & Toth, M. D. (2013). *Teacher evaluation that makes a difference: A new model for teacher growth and student achievement*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Medley, D. & Coker, H. (1987). The accuracy of principals' judgments of teacher performance. *Journal of Educational Research*, 80(4), 242-247.
- McQuarre, F. O. J., & Wood, F. H. (1991). Supervision, staff development and evaluation connections. *Theory in to Practice*, 30, 91-96.
- McGuinn, P. (2012). Stimulating reform: Race to the top, competitive grants and the Obama education agenda. *Educational Policy*, 26(1), 136-159. doi: 10.1177/0895904811425911.
- Mills, G. (2003). *Action research: A guide for the teacher researcher* (2nd ed.). Columbus, OH: Merrill Prentice Hall.
- Minnear-Peplinski, R. M. (2009). *Principals' and teachers' perceptions of teacher supervision* (Doctoral dissertation, University of Nevada Las Vegas). Retrieved from <http://digitalscholarship.unlv.edu/cgi/viewcontent.cgi?article=1056&context=thesedsdissertations>

- Mooney, J. (2012). School districts comparison shop for teacher evaluation systems. Retrieved from <http://www.njspotlight.com/stories/12/10/17/school-districts-comparison-shop-for-teacher-evaluation-systems/>
- New Jersey Department of Education. (2014). *Guide to the TeachNJ (2012) Act*. Retrieved from <http://www.state.nj.us/education/AchieveNJ/intro/TeachNJGuide.pdf>
- New Jersey Department of Education. (2014). *NJ ASK parent, student, and teacher information guide*. Retrieved from http://www.state.nj.us/education/assessment/es/njask_info_guide.pdf
- NJPSA, NJASA, & NJEA. (2014). *Voices from the Field: A Collaborative Survey on Educator Evaluation Implementation*. Retrieved from <http://www.njpsa.org/documents/pdf/VoicesfromtheField-DigitalFolder.pdf>.
- NJSBA (2019, March 26). NJDOE Announces Connected Action Roadmap. *School Board Notes*. Retrieved from <https://www.njsba.org/news-publications/school-board-notes/march-26-2019-vol-xlii-no-33/car-talk-njdoe-announces-connected-action-roadmap-car-regional-information-sessions/>
- No Child Left Behind Act. Pub. K. 107-110 (2001). Retrieved from <http://www2.ed.gov/policy/elsec/leg/esea02/107-110.pdf>
- Papay, J.P. (2012). Refocusing the debate: Assessing the purposes and tools of teacher evaluation. *Harvard Educational Review*, 82(1), 123-141.
- PARCC . (2016). <http://www.parcconline.org/>
- Patton, MQ. (1999). "Enhancing the quality and credibility of qualitative analysis." *HSR: Health Services Research*. 34 (5) Part II. pp. 1189-1208.
- Patton, M.Q. (2002). *Qualitative evaluation and research methods* (3rd ed.). Thousand Oaks, CA: Sage.
- Porter, A., McMaken, J., Hwang, J., & Yang, R. (2011). Common core standards: The new U.S. intended curriculum. *Educational Researcher*, 40(3), 103-116.
- Race to the Top. (2009). <https://www2.ed.gov/programs/racetothetop/index.html>
- Ramirez, A. (2010, December). Merit pay misfires. *Educational Leadership*, 68(4), 55-58.
- Robinson, G. (2009). Principals' perceptions regarding the role of the professional development and appraisal system in teacher supervision in Texas. Retrieved from ProQuest Digital Dissertations. (AAT 3358085)

- Rubin, H., & Rubin, I. (2012). *Qualitative interviewing: The art of hearing data* (3rd ed.). Los Angeles, CA: Sage.
- Rucinski, M., & Diersing, C. (2014, May 14). America's teacher evaluation system revolution. *Harvard Political Review*. Retrieved from <http://harvardpolitics.com/united-states/americas-teacher-evaluation-system-revolution/>
- Rundquist, J. (2009, October 7). Issues facing N.J. public schools, students tackled by gubernatorial candidates. Retrieved from: http://www.nj.com/news/index.ssf/2009/10/issues_facing_nj_public_school.html
- Saldana, J. (2013). *The coding manual for qualitative researchers* (2nd ed.). Thousand Oaks, CA: Sage.
- Schoonenboom, J., & Johnson, R. B. (2017). How to Construct a Mixed Methods Research Design. *Kolner Zeitschrift für Soziologie und Sozialpsychologie*, 69(Suppl 2), 107–131. doi:10.1007/s11577-017-0454-1
- Shields, C. (2010). Transformative leadership: Working for equity in diverse contexts. *Educational Administration Quarterly*, 46, 558-589. (PDF)
- Sheppard, J. D. (2013). *Perceptions of teachers and administrators regarding the teacher evaluation process* (Doctoral dissertation, Georgia Southern University). Retrieved from <http://digitalcommons.georgiasouthern.edu/cgi/viewcontent.cgi?article=1856&context=etd>
- Spring 2019 State Assessment Name, Length and Time Changes [Broadcast]. (2018, October 5). Retrieved from https://homeroom5.doe.state.nj.us/broadcasts/2018/OCT/05/19026/Changes to Spring 2019 State Assessment Formats and Policies_Jan 11.pdf
- Stephey, M. J. (2008, November 17). A brief history of tenure. *Time Magazine*. Retrieved from <http://content.time.com/time/nation/article/0,8599,1859505,00.html>
- Stronge & Associates. (2015). *Stronge+ teacher effectiveness performance evaluation system handbook* [Handbook]. VA: Stronge & Associates Educational Consulting, LLC.
- Stronge, J. (Ed.). (2006). *Evaluating teaching* (2nd ed.). Thousand Oaks, CA: Corwin Press.
- Superfine, B.M., Gottlieb, J.J., & Smylie, M.A. (2012). The expanding federal role in teacher workforce policy. *Educational Policy*, 26(1), 58-78. doi: 10.1177/0895904811435722.

- Tashakkori, A. & Teddlie, C. (Eds). (2003). *Handbook of mixed methods in social and behavioral research*. Thousand Oaks, CA: Sage.
- Taylor, E. , & Tyler, J. (2012). Can teacher evaluation improve teaching? Evidence of systematic growth in the effectiveness of midcareer teachers. *Education Next*. Retrieved January 31, 2013
- TeachNJ (2012) Act, NJ Assembly, 2012, Chapter 26 (NJ Stat. 2012).
- Towe, P. B. (2012). *Investigation of the role of a teacher evaluation system and its influence on teacher practice and professional growth in four urban high schools* (Doctoral dissertation, Seton Hall University). Retrieved from <http://scholarship.shu.edu/cgi/viewcontent.cgi?article=2792&context=dissertations>
- Vinovskis, M. A. (1992). Schooling and poor children in 19th century America. *American Behavioral Scientist*, 35(3), 313-331. <http://dx.doi.org/10.1177/0002764292035003008>
- Wacha, M. (2013). *Teacher perceptions of the teacher evaluation process* (Master's thesis). Retrieved from <http://search.proquest.com.ezproxy.rowan.edu/pqdtglobal/docview/1352156360/49BD9F5F52E24779PQ/1?accountid=13605>
- Winslow, R. A. (2015). *Administrative feedback following classroom observations as part of a Danielson based teacher evaluation system: Teacher and administrator perceptions* (Doctoral dissertation). Retrieved from <http://search.proquest.com.ezproxy.rowan.edu/pqdtglobal/docview/1690904225/17EC8D7C794340C7PQ/1?accountid=13605>
- Zepeda, S. J. (2007). *Instructional supervision: Applying tools and concepts* (2nd ed.). Larchmont, NY: Eye on Education.
- Zepeda, S. J. (2017). *Instructional supervision: Applying tools and concepts* (3rd ed.). New York, NY: Routledge.

Appendix A

Survey Questions

1. Including the current year, how many years have you taught in City Public Schools?
 - a. 1 year
 - b. 2 to 5 years
 - c. 6 to 10 years
 - d. 11 to 15 years
 - e. 16 or more years
2. Select the answer that best describes your current teaching assignment
 - a. Pre-K- K
 - b. Grades 1 to 4
 - c. Grades 5 to 8
 - d. Grades 9 to 12
 - e. K-12
3. Please rate the overall quality of the evaluation process, including observations.
(1= very poor quality, 5 = very high quality)
4. Please rate the overall impact of the Stronge TEPES on your professional practices. (1= no impact, 5= strong impact)
5. Please rate the overall quality of the implementation of the Stronge TEPES in City Public Schools (1= very poor quality, 5 = very high quality)

Using the following 5-point Likert scale, please answer the following questions:

- 1- Strongly disagree
 - 2- Disagree
 - 3- Neither agree nor disagree
 - 4- Agree
 - 5- Strongly agree
6. The implementation of the Stronge TEPES has the ability to create positive change in the City Public Schools.

7. The written feedback that I was provided after formal observations helped me to identify strengths in my teaching.
8. The written feedback that I was provided after formal observations helped me to identify areas needing improvement in my teaching.
9. The written feedback that I received after formal observations was valuable.
10. The written feedback that I received after formal observations was meaningful in terms of identifying teaching practices I used.
11. The written feedback that I received after formal observations led to learning about teaching on my part.
12. The written feedback that I received after formal observations has had a positive influence on my teaching practices.
13. The written feedback that I received after formal observations caused me to reflect on my teaching.
14. The written feedback that I received after formal observations has led or will lead to positive changes in my teaching practices.
15. My annual evaluation leads to getting professional development and support for my improvement.
16. I think teacher evaluation improves the learning of my students.
17. Please indicate if you would be willing to participate in a follow up interview
 - a. Yes
 - b. No
18. If you answered yes to question 17, please write your name below:

Appendix B

Interview Protocol

1. What is your current position in City Public Schools?
2. How long have you been in this position?
3. Describe your perception of the implementation of the Stronge TEPES in the City Public Schools.
4. How did the implementation create a change in the district, either positively or negatively?
5. How well prepared did you feel for the implementation of the Stronge TEPES?
6. How well prepared did you feel the district was for the implementation of the Stronge TEPES?
7. Describe the quality of feedback you are receiving on either classroom observations or summative evaluations?
8. How, if at all, has the feedback affected your instruction?
9. How has the implementation of the Stronge TEPES changed the professional growth opportunities for teachers in the district?
10. In what ways has the supervision of teachers changed?